UNDERSTANDING CHICKEN BG GENES AT THE RNA AND PROTEIN LEVELS

Lei Chen

Darwin College

UNIVERSITY OF CAMBRIDGE

Department of Pathology

This dissertation is submitted for the degree of
Doctor of Philosophy

September 2018

Understanding chicken BG genes at the RNA and protein levels

Lei Chen

The chicken BG system is a highly polymorphic and polygenic multigene family encoding type I transmembrane proteins, with butryophilins as homologues in mammals, some of which are crucial in T cell regulation. There are three genomic locations where BG genes are found: one singleton BG gene (BG0) on chromosome 2, another singleton gene (BG1) in BF-BL region (the so-called minimal essential chicken MHC) on chromosome 16, and many BG genes arranged tandemly in the BG region just outside the MHC. BG genes in BG region have copy number variation between different chicken haplotypes, so it has been unclear which BG genes are alleles, as very little sequence information has been available for haplotypes other than B12, the best characterized one. Also, the functions of chicken BG genes have been a mystery for half a century, although there is evidence for cytoskeletal regulation and for viral disease resistance. Therefore, the aim of the research was to develop new procedures and reagents to understand the BG system.

A novel PCR protocol was established to overcome the difficulty of amplifying full length polymorphic BG transcripts, and then was applied to systematically examine the BG cDNA sequences from T cells and B cells of four different chicken haplotypes. In total 23 BG genes were found, most with alternative splicing isoforms; most strikingly, the transcripts potentially encoding soluble BG proteins were only seen in B cells, indicating functional differences of the same gene in T and B cells. By comparing the dominantly expressed BG genes as ‘functional alleles’ in these cells, only the cytoplasmic tail region is clearly seen to be under selection, based on the overwhelming preponderance of non-synonymous changes. With many other unexpected findings discovered in this project, a clearer picture of chicken BG genes is presented, and more questions were raised for future study.

In order to explore BG functions and further characterize BG proteins, fourteen stable cell lines were developed expressing fusion proteins of the Ig-V domains of the 14 BG genes from the B12 haplotype chicken with the human IgG1 Fc fragment. These BG-Fc fusion proteins were used in sandwich enzyme-linked immunosorbent assays (ELISAs) to screen 290 BG monoclonal antibody (mAb) tissue culture supernatants, and these BG mAbs were further characterized for specificity by western blot using BG-Fc fusion proteins. These solid tools (BG-Fc fusion proteins and BG mAbs) provide the basis to further understand chicken BG functions and answer other interesting questions.

Acknowledgements

I would like to take this opportunity to sincerely thank my supervisor, Prof. Jim Kaufman, for his patience and kindness during the past four years training me to be a better scientist; for his time and efforts of setting up a model to show that hard working matters as much as passion and talent; for this encouragement and jokes to allow me to look at the world from a different angle. I wish Jim good health and spirit, enjoy work and life, and can be proud of having me as his PhD student in the future.

I would like to thank all members in Kaufman's lab who help me so much and are just so wonderful to work with. Mr. Dan Wise supported me as much as he can during our PhD journey. Dr. Alisa Martin-Lopez introduced me many interesting aspects of life in UK. Dr. Michaela Fakiola accompanied me for many nights working in the lab, Dr. Clive Tregaskes kindly answered almost all my questions no matter how naïve they were. Dr. Samer Halabi gave me many suggestions during the late stage of my study and I enjoyed many nice conversations. Dr. Hassnae Afraache helped me with bioinformatics questions and shared her valuable experience with me in writing thesis. Dr. Kahina Meziane, Dr. Andrew Chan, Mr. Michael Harrison, Ms. Becky Martin, Ms. Flic Coulter, Ms. Eve Doran, Mr. Amar Sharma and Mr. Sam Hilton all have been so friendly.

I would also like to thank some of the members in the Department of Pathology. Prof. John Trowsdale and Dr. Adrian Kelly kindly gave me many suggestions during my first year viva. Special thanks to my mentor, Dr. Brian Ferguson, who kindly supported me, listened to my stories and gave me many constructive suggestions. Also very special thanks to my good friend Dr. Chiwen Chang, who kindly helped out with some experiments and shared many funny stories while having meals in Chinese and Thai restaurants.

I am grateful to Dr. John Chattaway and Ms. Valeria Radjabova who helped me a lot during the initial stage of my PhD study; Dr. Colin Butter and Ms. Karen Staines kindly provided samples for my study; Dr. Michaela Fakiola taught me how to prepare samples for sorting; Mr. Nigel Miller helped me with sorting; Dr. Aimee Parker taught me how to do section and staining.

Finally, I would like to thank my family for their caring and being supportive, and I dedicate this thesis to them.

Preface

I hereby declare that

- This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the acknowledgements and specified in the text.
- It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University of similar institution except as declared in the Preface and specified in the text.
- It does not exceed the prescribed word limit for the relevant Degree Committee.

..... Lei Chen .....

.....

2019-01-18

.....

Contents

Table of Contents

Appendixes

List of figures

List of tables

Abbreviations

AEBSF	- 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride
AIV	- avian influenza virus
ATP	- adenosine triphosphate
BAC	- bacterial artificial chromosome
BSA	- bovine serum albumin
bp	- base pairs
BLAST	- basic local alignment search tool
BLAT	- BLAST-like alignment tool
BTN	- butyrophilin
BTNL	- butyrophilin-like
cDNA	- complementary DNA
CD	- cluster of differentiation
CDR	- complementarity determining regions
CDS	- coding sequence
CNV	- copy number variation
CTLA-4	- cytotoxic T-lymphocyte-associated antigen 4
Da	- Dalton
DAP-10	- DNAX adaptor protein-10
DAP-12	- DNAX adaptor protein-12
DEPC	- diethyl pyrocarbonate
DMEM	- dulbeco's modified eagle's medium
DMSO	- dimethyl sulfoxide
DNA	- deoxyribonucleic acid
dNTP	- deoxynucleotide triphosphate
dN/dS	- non-synonymous/synonymous SNP ratio
EBI	- European Bioinformatics Institute
EDTA	- ethylene diamine tetra-acetic acid
ELISA	- enzyme-linked immunosorbent assay
ER	- endoplasmic reticulum
FACS	- fluorescence-activated cell sorting
FBS	- fetal bovine serum

Fiber FISH	- fiber fluorescent in situ hybridization
FISH	- fluorescent in situ hybridization
FITC	- fluorescein isothiocyanate
GAPDH	- glyceraldehyde 3-phosphate dehydrogenase
HRP	- horse radish peroxidase
IBDV	- infectious bursal disease virus
IBV	- infectious bronchitis virus
Ig	- immunoglobulin
Ig-V	- Ig-like V
ITAM	- immunoreceptor tyrosine-based activation motif
ITIM	- immunoreceptor tyrosine-based inhibitory motif
Kb	- kilobase
Kd	- kilodalton
KIR	- killer cell immunoglobulin receptor
mAb	- monoclonal antibody
MDV	- Marek's disease virus
MHC	- major histocompatibility complex
mg, μg, ng	- miligram, microgram, nanogram
ml, μl	- mililiter, microliter
mM, μM	- milimolar, micromolar
mRNA	- messenger RNA
MOG	- Myelin Oligodendrocyte Glycoprotein
NCBI	-National Centre for Biotechnology Information
NJ	- neighbour joining
NK	- natural killer cells
PAGE	- polyacrylamide gel electrophoresis
PBL	- peripheral blood lymphocytes
PBS	- phosphate buffered saline
PCR	- polymerase chain reaction
PDB	- protein data bank
PE	- phycoerythrin
REV	- reticuloendotheliosis virus
RFLP	- restriction fragment length polymorphism

RJF	- red junglefowl
Rfp-Y	- restriction fragment pattern Y
RNA	- ribonucleic acid
RSV	- Rous sarcoma virus
RT-PCR	- reverse transcriptase polymerase chain reaction
SDS	- sodium dodecyl sulphate
SDS-PAGE	- sodium dodecyl sulphate polyacrylamide gel electrophoresis
SNP	- single nucleotide polymorphism
TAE	- Tris acetate
TAP	- transporter associated with antigen processing
TBS	- Tris buffered saline
TBST	- Tris buffer saline with tween 20
TCR	- T cell receptor
TE	- Tris EDTA
TEMED	- tetramethylethylenediamine
TM	- transmembrane region
Tm	- the melting temperature of the primer or oligo
TRIM	-Tripartite motif containing
U	- units
UTR	- untranslated region
WGS	- whole genome shotgun
^{0}C	- degrees celsius

Chapter 1

General introduction

1.1 The chicken B locus

1.1.1 The chicken B locus contains three loci: BF, BL and BG

The chicken B locus was first discovered in 1950 as encoding a red blood group antigen (Briles *et al*., 1950) which after a decade was revealed as a histocompatibility system during skin grafting experiments (Schierman *et al*., 1961). Several approaches demonstrated that the chicken B locus contain three loci: BF, BL and BG (Ziegler *et al*., 1976; Hála et al. 1976; Pink et al. 1977). The name of 'major histocompatibility (B) complex of chicken' was first published in 1977 with a picture showing chicken B locus containing two regions, one for BG locus, and the other one for BF and BL loci responding for the tolerance or rejection during skin grafting (Pink et al. 1977). Later, BF and BL loci were confirmed to encode chicken MHC class I and class II molecules, respectively (Bourlet *et al.*, 1988; Guillemot *et al.*, 1988; Kaufman *et al.*, 1992). It's worth mentioning that MHC is well known for its polymorphism though, when the chicken B locus was found, the serological alloreactivity was due to the polymorphic BG antigens on erythrocytes (Pink *et al.*, 1977; Vilhelmová *et al.*, 1977; Simonsen *et al.*, 1980).

The genomic mapping of chicken B locus was reported since 1988, with enormous efforts from many groups (Guillemot *et al.*, 1988; Chaussé *et al.*, 1989; Pickel *et al*., 1990; Kaufman *et al*., 1991; Moiler *et al.*, 1991; Miller *et al.*, 1994; Miller *et al.*, 1994; Kaufman *et al*., 1995); some milestone publications helped to understand the genomic locations of BF, BL and BG, as well as the compositions of molecules from each locus. Kaufman et al. first drew the whole picture for BF/BL region and proposed 'the chicken B locus is a minimal essential major histocompatibility complex'. This 92-kilobase region of the B locus is roughly 20-fold smaller than human MHC and contains a minimal essential set of 19 MHC genes, but the organizations differ from their counterparts in mammals (Kaufman *et al.*, 1999). Later, a region called tripartite motif (TRIM) encoding several TRIM genes was identified between BF/BL and BG regions (Régnier *et al.*, 2003; Maruoka *et al.*, 2005; Ruby *et al.*, 2005; Salomonsen *et al.*, 2005). Also, two class I-like genes, CD1.1 and CD1.2, were found located adjacent to the MHC region (Maruoka *et al*., 2005). More recently, Salomonsen et al. determined the genomic sequence of the whole BG locus from a cosmid library constructed from the genomic DNA of a CB congenic chicken line (B12 haplotype), and found 12 BG genes arranged in the same orientation with a total of 99,274 bp in length (Salomonsen *et al.*,

2014). Thus far, a full picture of chicken B locus was presented in figure 1.1 based on B12 haplotype chicken (Kaufman, 2014). Overall, the chicken B locus fully sequenced so far is about 250 kb in a red junglefowl with BQ haplotype according to the whole genome shotgun gallus gallus 5.0 in Ensembl, but varies between different haplotypes mainly due to the BG region which has undergone expansion and contraction (Hillier *et al.*, 2004; Salomonsen *et al.*, 2014). There are three regions: the BF/BL region encodes MHC genes, BG region encodes the multigene family of BG genes, and in between a TRIM region encoding several TRIM genes.

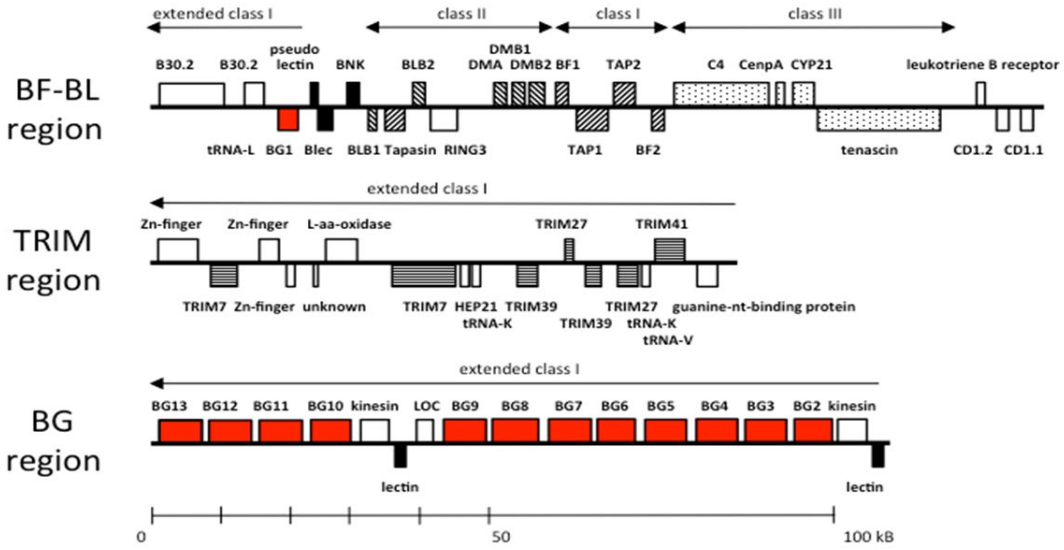


Figure 1.1 Genomic mapping of chicken B locus with currently sequenced genes based on B12 haplotype chicken (Kaufman, 2014)**.**

The chicken B locus contains three loci: the BF/BL loci, upstream of which there is a TRIM region, and the BG locus which is upstream of the TRIM region. Genes currently been identified and sequenced are represented by boxes with gene names either above or below these boxes.

1.1.2 Chicken B locus is much simpler and more compact comparing to human HLA

The chicken B locus (or B complex), the chicken MHC, is much simpler and more compact than the MHC of other mammals but differs in genomic organization (Kaufman *et al.*, 1999; Kaufman, 2008, 2014, 2018). Figure 1.2, the comparison between genes and regions in chicken B complex and human HLA, adapted from the most recent publication, best summarized the similarities and differences. First of all, human HLA is large and complex, with more than 200 genes spanning over 3.6 Mb (Beck *et al*., 2000; Chaves *et al.*, 2010; Chan, 2014); in contrast, chicken B complex is small and compact, about 242 kb in length based on current red junglefowl BAC clones (Shiina *et al.*, 2007) with only 19 genes in the BF/BL region, sixteen more in TRIM region and additional multigene family BG genes in BG region (Kaufman, 2008; Salomonsen *et al.*, 2014). Secondly, the genomic organizations are different. Human HLA class I and class II regions are separated by class III region, with so-called extended class I and extended class II regions on the end of class I and class II regions, respectively, and the TRIM region is located in between the class I and extended class I (Beck *et al*., 2000; Kaufman, 2018b). Chicken BF (class I) and BL (class II) genes are next to each other, with a minimal class III region to the end of the BF region. Downstream of the class III genes, there is a Y region [short for restriction fragment polymorphism (Rfp) - Y region] genetically separated by GC-rich region to the B locus and encoding at least one non-classical class I gene, one non-classical class II B and lectin-like genes (Miller *et al.*, 2004; Kaufman, 2014); on the other end, upstream of the class I, there is the so-called extended class I, upstream of which is the TRIM region, and then the BG region (Kaufman, 2018b). Thirdly, chicken B complex has the minimal essential MHC, meaning a lot of genes found in human HLA are not detected in chicken B complex; however, there are some unique genes found in chicken B complex but not seen in human HLA, for example the large number of BG genes in BG region as well as one BG gene found in BF/BL region.

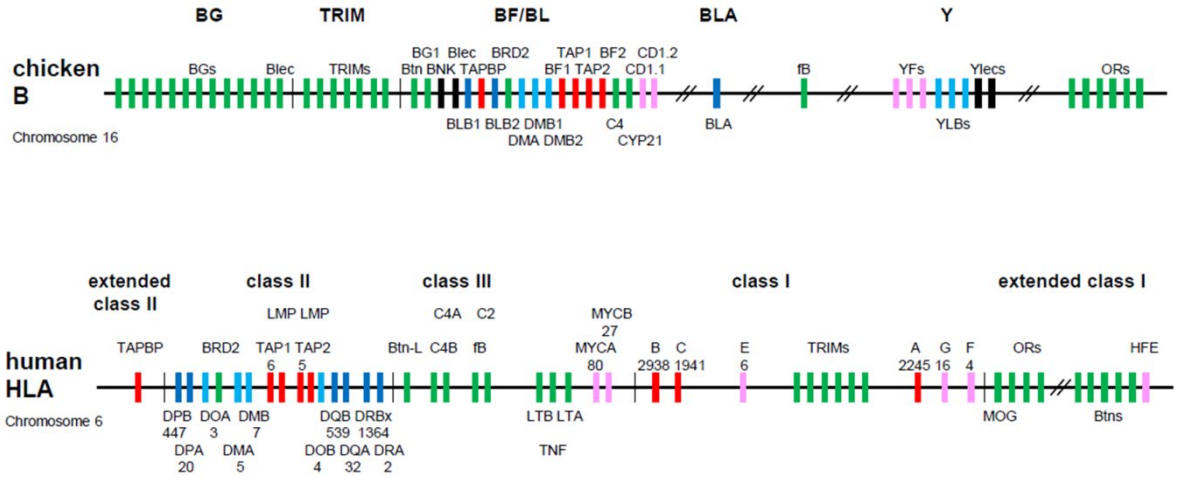


Figure 1.2 Comparison of chicken B complex and human HLA (Kaufman, 2018b)**.** Figure adapted from Kaufman 2018, a complete representation of the most important genes and regions in chicken and human MHC. Solid horizontal lines indicate MHC class I and class II genes, with red for classical class I, pink for non-classical class I, dark blue for classical class II, light blue for non-classical class II, and green for complement component genes. Dotted vertical lines indicate antigen processing and peptide loading genes, with red for class I system, and blue for class II system. Thin vertical lines indicate separation of defined regions (Kaufman, 2018b). The diagrams in this figure are not to scale.

1.1.3 BF/BL region

The BF/BL region, also called 'classical MHC' or the 'core MHC' (Shiina *et al.*, 2007), has been intensively studied over the past decades due to the importance of the associations with important infectious diseases that influence poultry industry heavily, as well as the research impact for providing insights into the evolution of acquired immune system for other species (Kaufman, 2018a, 2018b; Miller *et al.*, 2016). Unlike human HLA, the BF/BL region only encodes the essential MHC genes (Figure 1.3) (Kaufman, 2018a). Two classical MHC class I molecules, the BF1 and BF2 (Koch *et al.*, 1870; Kaufman *et al.*, 1999), flanking two TAP molecules, the TAP1 and TAP2, are located in the BF region (class I region). Two classical MHC class II B genes, the BLB1 and BLB2 lying on the two sides of Tapasin gene, together with a few other genes participating in peptide loading in the MHC class II system, DMA, DMB1 and DMB2, are located in BL region (class II region). Several other genes, C4, CenpA, CYP21, Tenascin, leukotriene B receptor, CD1.1, CD1.2 etc., are located in the class III region.

BF and BL are responsible for classical peptide antigen presentation (Kaufman, 2018b; Miller *et al.*, 2016). In the MHC class I system, both BF1 and BF2 encode classical MHC class I molecules, although BF2 is found dominantly expressed (Livant *et al.*, 2001; Wallny *et al.*, 2006; Shaw *et al.*, 2007) and coevolve with TAP2 (Walker *et al.*, 2011; Kaufman, 2015). Evidence supports the idea that BF2 is the major MHC class I molecule presenting peptides (Koch *et al.*, 2007; Wallny *et al.*, 2006; Butter *et al.*, 2013), while BF1 has gone through various deletion, insertion and rearrangements (Shaw *et al.*, 2007; Neill *et al.*, 2009; Miller *et al.*, 2018), and some studies suggested that BF1 might be ligand for receptors on NK cells (Ewald *et al.*, 2004; Livant *et al.*, 2004; Kim *et al.*, 2018; Miller *et al.*, 2018). The MHC class II system is more complicated compared to MHC class I system. The BLB1 and BLB2 genes (both encoding class II β chains) together with BLA gene (encoding class II α chain but mapped 5 cM away from the BF/BL region) encode the MHC class II molecules (Salomonsen *et al.*, 2003). Similar to the class I system, one class II gene, BLB2, is expressed at high level and probably co-evolves with DMB2 (Kaufman, 2008; Parker, 2013), while BLB1 most likely coevolves with DMB1 and is expressed most strongly in some intestinal tissues and spleen (Parker and Kaufman, 2017).

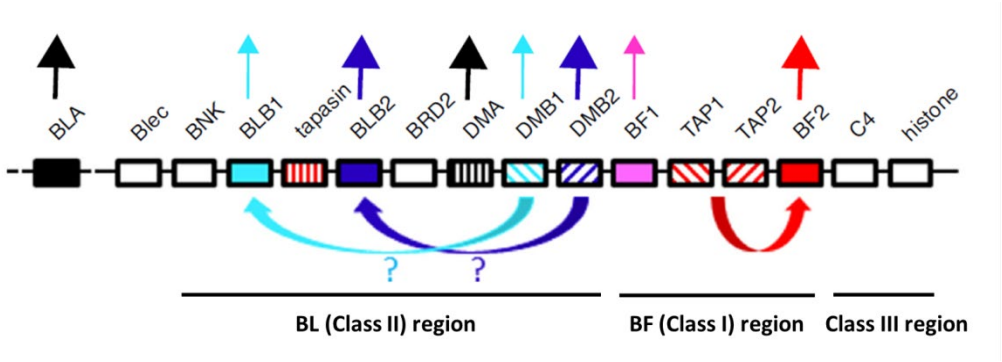


Figure 1.3 The BF/BL region of chicken B locus (adapted from Parker and Kaufman, 2017).

MHC genes in chicken BF/BL region are indicated as boxes with gene names above. In the BF (Class I) region, the polymorphic TAP genes have coevolved with BF2 which is dominantly expressed compared to BF1. In the BL (Class II) region, two class II B genes, BLB1 and BLB2 most likely coevolve with peptide loading chaperone genes DMB1 and DMB2, respectively, and BLB2-DMB2 are highly expressed in many tissues and cell lines, while BLB1-DMB1 are poorly expressed except in the intestine. The arrows pointing up indicate level of expression in blood and spleen; genes in red and pink are from class I system, in dark blue, light blue and black are from class II system; solid colours indicate classical genes and striped colours indicate genes involved in peptide loading (Parker and Kaufman, 2017).

1.1.4 Chicken B haplotype

The chicken B locus (B complex) has been well recognized to confer resistance or susceptibility to particular pathogens (Briles *et al.*, 1977; Collins *et al.*, 1977; Schierman *et al.*, 1977; Lamont *et al.*, 1987; Charleston *et al.*, 1998; Wald, 2005; Goto *et al.*, 2009; Schou *et al.*, 2010; Norup *et al.*, 2011; Schulten *et al.*, 2009; Cotter *et al.*, 1998; Dawes *et al.*, 2014). Breeding in industry and in chicken research lab have selected for various chicken B haplotypes mainly in the white lenghorn breeds (Simonsen *et al.*, 1982).

The chicken B haplotype includes information for BF, BL and BG alleles. The three loci in chicken B locus are not equally linked. BF and BL loci are closely linked without recombination observed by using haemagglutination or restriction fragment length polymorphism (RFLP) method (Hala *et al.*, 1979; Simonsen *et al.*, 1982; Skjødt *et al.*, 1985; Hála *et al.*, 1988). BG locus is linked to BF/BL region, but with a low and measurable rate of recombination (Briles *et al*., 1977; Hála *et al*., 1981; Koch *et al.*, 1983; Chattaway, 2013). An example to show recombination between BG region and BF/BL region is the R^1 haplotype which carried the BG alleles from CC strain and BF/BL alleles from the CB strain (Pink *et al.*, 1977). As some of later work has mapped certain disease resistance specifically to the BF/BL region, the use of MHC haplotype referring to BF/BL haplotype sometimes caused confusion with chicken B haplotype. Therefore, one should be very careful when using such nomenclature.

In order to clarify and unify the nomenclature of chicken haplotypes, a workshop was held in 1981 (Briles *et al.*, 1982). First of all, a total of 27 standard chicken strains were designated of their B haplotypes (from B^1 to B^{29} without B^{16} nor B^{20}), and all these standard B haplotype have their individual BF, BL and BG alleles using the same haplotype name. For example, B^1 haplotype has BF^1, BL^1 and BG^1; B^{29} haplotype has BF^{29}, BL^{29} and BG^{29}. Second, eight recombinant B haplotypes (with recombination between BG region and BF/BL region) were proposed to have names which helped to understand the BF/BL alleles but not BG alleles. Take R^1 haplotype mentioned above as an example, R^1 has the BG 4 allele and BF 12, BL 12 alleles, and thus was given the new name B^{12r1} with the number '12' standing for BF/BL allele and '1' for the first recombination event observed. However, all the typing to determine the chicken B haplotypes were done based on serological interactions using limited resources available at that time: some monoclonal antibodies to BG antigens but without knowing whether they were specific for particular BG antigens or broadly crossed-reactive

(Longenecker *et al.*, 1979; Briles *et al.*, 1982), alloantisera specific for BF and BG antigens (Hála *et al.*, 1976b), and alloantisera potentially detecting both BF and BG antigens (Hála *et al.*, 1976b; Longenecker *et al.*, 1979; Briles *et al.*, 1982).

With the development of sequencing techniques, the knowledge about chicken B haplotypes is enriched by the sequence information from BF and BL alleles. If we search 'chicken MHC class I' and 'chicken MHC class II' in 'Nucleotide' under NCBI website (https://www.ncbi.nlm.nih.gov/nuccore), there are almost two thousand nucleotide sequence records found for chicken MHC class I and class II molecules. Under the general guidance of 1982 nomenclature, a workshop discussing annotating those allelic sequences for MHC genes (not BG genes) was held (Miller *et al.*, 2004). More recently, an online database called 'IPD-MHC CHICKEN allele search' (https://www.ebi.ac.uk/ipd/mhc/group/CHICKEN/allele?page=1), under the 'Immuno Polymorphism Database' (IPD), was formed to help gathering and organizing all the available BF1 and BF2 genes with specified allelic information. Unknown MHC class I sequences can be blasted to search the database to understand their relationships with well-defined ones in the database within seconds. However, much less sequencing information is available for BG genes.

1.2 Chicken BG genes

The chicken BG locus encodes a multigene family of BG genes arranged in the same orientation, upstream of the TRIM region and BF/BL region; there are another two singleton BG genes with one found in BF/BL region on chromosome 16, and the other one on chromosome 2 (Salomonsen *et al.*, 2014). As mentioned already, the chicken has a minimum essential MHC, with 92 kb encoding almost all the MHC molecules (Kaufman *et al.*, 1999); while BG locus is closely linked to BF/BL region with roughly the same size (99 kb in B12 haplotype). Therefore, it is most likely that BG genes take important roles in immune system. Unfortunately the research on BG genes is much slower than that on MHC, and two publications summarized almost all the knowledge that we knew about BG, with one review paper published in 1991 (Kaufman *et al.*, 1991) and a more recent research paper in 2014 (Salomonsen *et al.*, 2014).

1.2.1 BG gene and protein structure

BG genes have similar intron-exon structures, but they differ in size mainly due to the exons encoding cytoplasmic tail. As shown in figure 1.4, exon 1 encodes 5' UTR (varies in size in different BG genes) and the signal sequence (99 bp), exon 2 encodes most of the immunoglobulin variable-like domain (Ig-V domain) (342 bp), exon 3 encodes the transmembrane region (105 bp), many 21 nucleotide exons (quantity varies in different genes) with or without a few 18 or/and 24 nucleotide exons encode the cytoplasmic tail, and a final exon encodes the last part of the cytoplasmic tail and 3' UTR (variable sizes) (Chattaway, 2013; Salomonsen *et al.*, 2014).

BG genes encode type I transmembrane proteins. The extracellular domain of BG protein is an Ig-V domain, where three cysteines are conserved in all BG genes found so far. Two cysteines form the intra-domain disulfide bond and the other one is located in the equivalent of complementarity determining region 1 (CDR1) forming a disulfide bond between the two chains of a BG dimer. The transmembrane regions of all 14 BG genes found in B12 haplotype are divided into two groups, with one group having histidine and lysine near the N-terminus while another group having leucine and threonine instead. The intracellular α-helical cytoplasmic tail forms a so-called coiled coil by interacting with another chain of BG cytoplasmic tail.

BG proteins are found as dimers with both homodimers and heterodimers observed but without any glycosylation identified (Salomonsen *et al.*, 1987; Kaufman *et al.*, 1991). The natural BG proteins immunoprecipitated with alloantisera or monoclonal antibodies (mAbs) by several groups showed that the bands of BG antigens normally ranged from 75 to 98 kD under non-reduced condition. However, under reduced conditions, either one band with roughly half size or two bands with different sizes have been observed, indicating that both homodimer and heterodimer can exist (Miller *et al.*, 1984; Salomonsen *et al.*, 1987; Kline *et al.*, 1988). Also, much larger sizes (135 and 160 kD) were detected under non-reduced gel suggesting higher multimers might be formed (Miller *et al.*, 1990).

However, it is still unclear whether the heterodimers are encoded by different BG genes or the same BG gene with the length variation due to the alternative splicing in the cytoplasmic tail (Kaufman *et al.*, 1991). Theoretical calculation of protein molecular weight based on amino acid sequence varies for BG proteins mainly due to the variable exon quantities and

compositions in the cytoplasmic tail regions. Take BG7 and BG8 from B12 haplotype as an example, the molecular weights for one chain of each are 43.31 kD and 46.86 kD respectively, according to online protein weight calculator tool (https://web.expasy.org/compute_pi). Therefore, it is possible that the heterodimer observed under the reduced condition were from two different BG proteins. However, alternative splicing in the cytoplasmic tails was detected from the cDNA sequences of limited studies on some BG genes (Kaufman *et al.*, 1991), therefore, it also might be possible that those bands with different sizes were actually from the same BG protein but with alternatively spliced cytoplasmic tails.

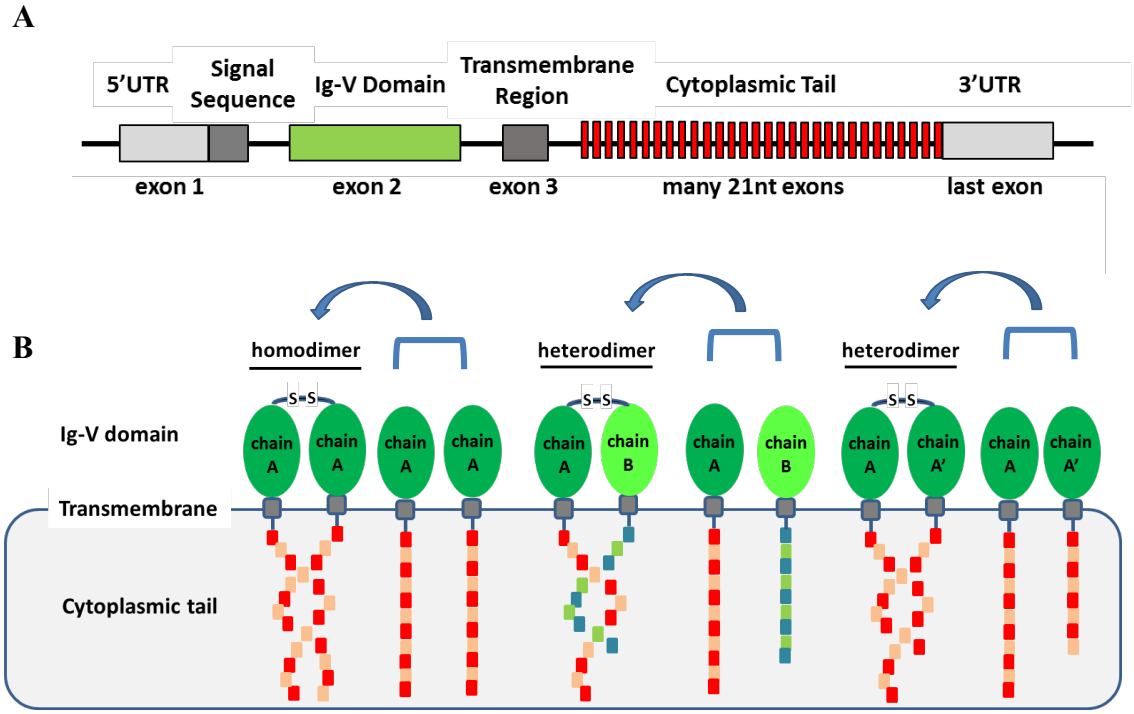


Figure 1.4 The BG gene and protein structures. A. The basic intron and exon structure of a BG gene which contains 5' terminal exon encoding a 5' UTR and a signal sequence, V domain exon, transmembrane region exon, coiled coil region exon and 3' terminal exon encoding 3' UTR. B. The model of BG proteins, which are disulfide-linked dimers of type 1 transmembrane protein. Both homodimer and heterodimer were observed, but it is unclear whether the heterodimer is formed by two different BG proteins or the same BG protein with different alternative splicing cytoplasmic tails. The coiled coil regions vary between different BG proteins. No N- or O-linked glycosylation site has been observed yet (Salomonsen *et al.*, 1987).

1.2.2 Genomic mappings of BG genes

There are three locations of BG genes mapped in the chicken genome (Salomonsen *et al.*, 2014). One singleton gene named BG0 is located on chromosome 2, another singleton BG gene called BG1 is located on chromosome 16 in the BF/BL region, and multiple BG genes in the BG region are also on chromosome 16.

The singleton BG gene, BG0, was first named as CTBG discovered from a caecal tonsil cDNA library by Prof. Jan Salomonsen and Dr. Olli Vainio (Dr. John Chattaway, personal communication), which was found identical in sequence to a partial cDNA clone sequenced by Miller et al. (Miller *et al.*, 1990). With the help from Prof. Jan Salomonsen, Dr. John Chattaway confirmed that BG0 is located on chromosome 2 of chicken genome where the genes surrounding BG0 are not related to immune system. Comparing the syntenic regions of human, mouse and *Xenopus*, no BG0 or BG-like molecule was found in other species. Dr. John Chattaway suggested that BG0 was most likely brought into this location by an insertion from a microchromosome during evolution (Figure 1.5) (Chattaway, 2013a).

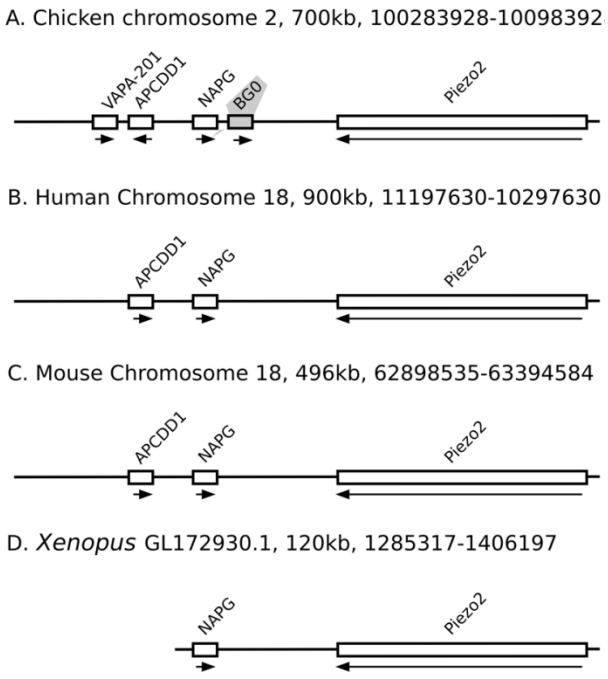


Figure 1.5 The genomic location of BG0 on chromosome 2 of chicken line CB compared to the equivalent locations on chromosomes of other species, figure copied from Chattaway, 2013.

The other singleton BG gene, BG1, is located in the BF/BL region which has been known for a long time (Figure 1.6). A gene named 8.5 was first identified in a cosmid by Southern blot using cDNA probe from a B cell tumor line (Guillemot *et al.*, 1988), and later it was confirmed to be a BG gene (Kaufman *et al.*, 1989). The name of 8.5 gene was later named BG1 (Shiina *et al.*, 2007), and since then BG1 has been used as a fixed name for the BG gene found located in the BF/BL region.

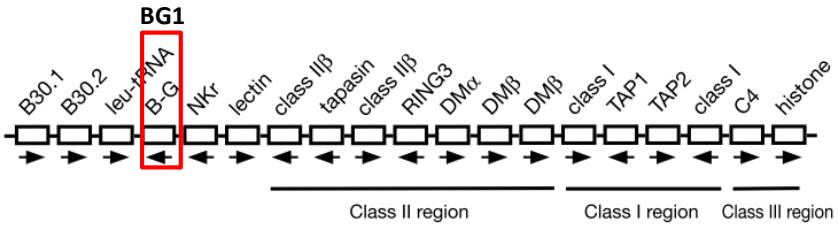


Figure 1.6 The genomic location of BG1 in the BF/BL region on chromosome 16 of chicken genome, figure adapted from Kaufman *et al.*, 1999.

The BG region containing multiple BG genes is about 80 kb away from the BF/BL region on chromosome 16 of chicken genome as estimated by few different groups (Ruby *et al.*, 2005; Shiina *et al.*, 2007). The most comprehensive mapping of BG genes in BG region is for CB line chicken (B12 haplotype): twelve BG genes are organized as tandem repeats with the same transcriptional orientation located upstream of the TRIM region on chromosome 16 (Figure 1.7) (Salomonsen *et al.,* 2014).

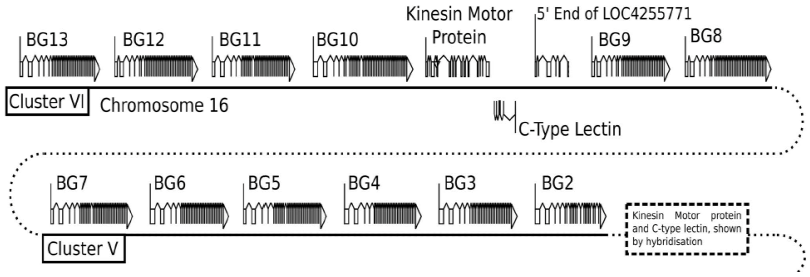


Figure 1.7 The location and orientation of 12 BG genes in BG region of CB line chicken (B12 haplotype) from Salomonsen *et al.*, 2014.

All 12 BG genes are presented with their introns, exons and intragenic regions to scale.

1.2.3 BG genes are highly polymorphic

BG antigens are known for their polymorphism. As mentioned previously, the chicken B locus was first discovered because of the polymorphic BG antigens rather than the MHC (Briles *et al.,* 1950; Pink *et al.*, 1977; Kaufman *et al.,* 1991); with the development of DNA sequencing technology, the polymorphism of BG genes are better understood at the molecular level.

In the early stages, serological typing played an important role to define BG molecules. Both monoclonal antibodies and alloantisera specific to BG antigens were used to identify BG proteins but mainly in erythrocytes (Hála *et al.,* 1976b; Longenecker *et al.,* 1979; Briles *et al.*, 1982; Salomonsen *et al.*, 1987; Kline *et al.*, 1988; Kaufman *et al.*, 1990; Miller *et al.*, 1991). Such work not only helped to understand the polymorphism of BG antigens in different chicken B haplotypes, but also reveals size differences of BG proteins may be due to alternative splicing in the cytoplasmic tails (Kaufman *et al.*, 1990).

Later, DNA sequence based typing methods were used to analyze BG genes and further support the fact that BG genes were polymorphic. For example, restriction fragment length polymorphism (RFLP) was used to distinguish BG genes from many different chicken B haplotypes (Miller *et al*., 1988), as well as to analyze BG patterns in commercial broilers from which many novel BG patterns were found which were different in each chicken company sampled, illustrating the great polymorphism of BG genes in commercial chicken lines (Li *et al*., 1997; Li *et al.*, 1999; Emara *et al.*, 2002). Another example is that 38 BG genotypes were found from 51 DNA samples of the Camperos chicken by Southern blot hybridization (Iglesias *et al.*, 2003).

Genomic sequencing recently fully characterized all 14 BG genes from CB line chicken (B12 haplotype), and none of the 14 BG genes is identical (Salomonsen *et al.*, 2014). Phylogenetic analyses based on nucleotide sequences of all 14 BG cDNA sequences annotated from genomic sequences reveal that there are two kinds of 5' UTRs, two types of cytoplasmic tails and two types of 3' UTRs. However, if looking at the phylogenetic trees built for each region of the gene, most BG genes clustered into the same group in one region are separated into different groups in other regions, which makes BG genes look like hybrid genes (Figure 1.8).

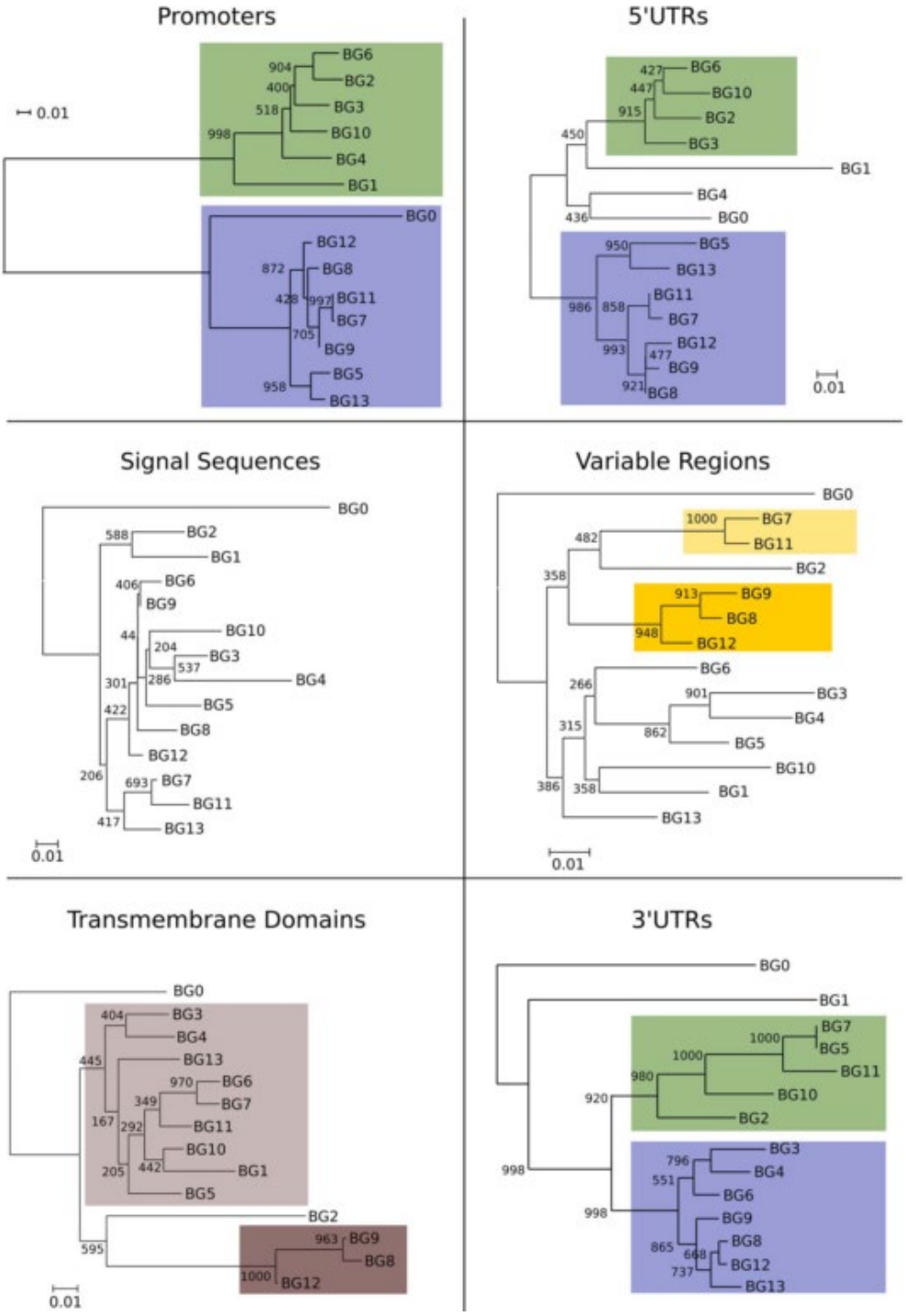


Figure 1.8 Phylogenetic analysis of 14 BG genes from B12 haplotype (figure copied from Salomonsen *et al.*, 2014). Phylogenetic trees are built on nucleotide sequence alignments using a Bayesian approach. BG genes are clustered differently in separate regions, which makes BG genes look like hybrid genes.

1.2.4 BG genes in BG region have copy number variation

There are at least two pieces of evidence showing that the BG genes in BG regions have copy number variation (CNV). One is from fibre-fluorescence *in situ* hybridization (fibre-FISH) assessing BG gene locations and compositions on chicken chromosomes of different haplotypes. Three cosmid probes (c4.5, cG43 and cG24) were made for CB line chicken (B12 haplotype), with c4.5 covering the BF/BL region, cG43 covering BG region from BG2 to BG7, and cG24 covering the rest part of BG region from BG8 to BG13. The genomic DNA from six different chicken haplotypes (B2, B4, B12, B15, B19 and B21) was tested by fibre-FISH using the three probes above. The results showed obvious differences in the size and composition of BG regions between different haplotypes, indicating the numbers of BG genes varies in the BG region of different haplotypes (Salomonsen *et al.*, 2014).

Another piece of evidence is from the comparison of BG genes between the B12 haplotype typed by Salomonsen et al. and the whole genome shot gun sequence (WGS) from a female red junglefowl, a BQ like haplotype (Hillier *et al.*, 2004). About 18 BG genes were considered to exist in the BG region of BQ haplotype compared to 12 BG genes in B12 haplotype, which also supports CNV in the BG region (Salomonsen *et al.*, 2014).

1.2.5 BG alleles

Since BG genes are so polymorphic, it is very important to know which BG genes are alleles in order to understand their functions. Until now, only the two singleton BG genes, BG0 and BG1, are known having alleles (Chattaway *et al.*, 2016), and there is no report regarding alleles of BG genes in the BG region.

1.2.5.1 BG0 alleles tested so far are virtually monomorphic

Nearly-full length cDNA sequences of BG0 were amplified and isolated from spleen samples of five different chicken haplotypes (B2, B4, B12, B15 and B21), and only a few single nucleotide polymorphisms (SNPs) were found when compared to the WGS BG0 sequence, suggesting BG0 might have a housekeeping function (Chattaway *et al.*, 2016). When amplifying BG0 cDNA sequences from different tissues and different chickens, a major transcript was present in all samples, with a few variant transcripts found due to alternative splicing either in the Ig-V domain or the cytoplasmic tail region (Chattaway *et al.*, 2016).

1.2.5.2 BG1 is polymorphic with point mutations throughout the whole sequence

The same work for BG0 above was done for BG1 and revealed that BG1 is highly polymorphic. The major transcripts of BG1 from all haplotypes show much variation throughout the whole sequence, with most concentrated in the Ig-V domain and 3' UTR region. Also, deletions and duplications of cytoplasmic tail exons were observed in these alleles (Chattaway *et al.*, 2016). Such results fit with genomic predictions by Hosomichi et al. on 14 chicken haplotypes (Hosomichi *et al*., 2008).

1.2.5.3 The attempt of identifying functional BG alleles failed

Prof. Jim Kaufman had designed a primer pair named UC74 and UC76, and collaborated with Prof. Jan Salomonsen to amplify the whole Ig-V domain of BG genes from a variety of tissues of different chicken haplotypes, trying to understand whether those tissue specific BG genes would reflect some common features in sequence as they may take same functions. However, this work failed to answer such question (Chattaway, 2013). As shown previously in figure 1.8, the Ig-V domain of BG genes are highly polymorphic, and unfortunately the BG genes clustered in the same group based on their Ig-V domain sequences were not necessarily from the same tissue. Therefore, in order to reveal the 'functional alleles' of BG genes, the whole cDNA sequence needed to be examined.

1.2.6 Tissue distribution

Comprehensive understanding of tissue distribution of BG genes is still lacking. Antigen-antibody interaction methods, e.g. immunoprecipitation, section staining, immunoblotting, flow cytometry etc., using BG monoclonal antibodies (mAbs) and/or anti-sera helped to prove the BG protein expressing in erythrocytes, thrombocytes, lymphocytes and lymphoid tissues (thymus, bursa, and caecal tonsil) (Miller *et al*., 1982; Miller *et al*., 1984; Kaufman *et al*., 1989; Salomonsen *et al.*, 1991). Also, BG mRNA was detected from the cell types and tissues above, also in small intestine and liver (Miller, 1990; Salomonsen *et al.*, 1991; Miller *et al.*, 1991).

Salomonsen et al. systematically assessed number of cDNA clones (as a proxy for the mRNA level) of all 14 BG genes of B12 haplotype in different cell types and tissues, and found striking expression patterns (Salomonsen *et al.*, 2014). As shown in figure 1.9, each BG gene has very specific cell and tissue expression. For example, in T cells, there are only two BG

genes found with BG7 strongly expressed and BG12 weakly expressed. Most significantly, all the BG genes found in haemopoietic cells labeled blue in figure 1.9 have their promoter and 5' UTR sequences clustered into the same group in phylogenetic trees in figure 1.8; on the contrary, the rest of the BG genes expressed in tissues labeled green have different promoter and 5' UTR clustered in another group in the phylogenetic tree, indicating the promoter and 5' UTR sequences might determine the tissue distribution of BG genes. However, such examination is based on mRNA level; further confirmation of BG protein expression should be done to confirm the exact tissue distribution.

cells/tissues		cluster VI						cluster V						cluster I	chr 2	other cDNA			
	gene name	BG13	BG12	BG11	BG10	BG9	BG8	BG7	BG6	BG5	BG4	BG3	BG2	BG1	BG0	I	II	No.	No.
	altern-	1	2	3	4	5	6	A	B	C	D	E	F	8.5	CTBG	B4x	B4y	seq	PCR
	ative	FB1	FB3	13A	13B	F8	B17	KCN	KCK	KM5	X8.7	X7.8	X23	FG	bg28				
	names			22E	zipper		F11			43F	14A		43A						
T cells		0	3	0	0	25	0	0	0	0	0	0	0/NT	0	0/+	13	1	42	8
B cells		3	1	0	0	18	2	16	0	0	0	0	0/NT	0	0/+	12	3	55	9
thrombocytes		5	5	4	0	0	5	4	0	6	1	0	0/NT	0	0/+	15	2	47	5
macrophage																			
	medium	0	0	0	0	14	0	16	1	2	0	0	0/NT	0	0/NT	0	0	33	4
	LPS	0	0	0	2	10	0	18	0	0	0	0	0/NT	0	0/NT	0	0	30	5
	INFg	0	0	0	0	18	0	16	0	0	0	0	0/NT	0	0/NT	0	0	34	5
	LPS + INFγ	0	0	0	0	2	0	8	0	0	0	0	0/NT	0	0/NT	0	0	10	3
dendritic cells		1	0	0	0	29	0	0	0	0	0	0	0/NT	0	0/NT	0	0	30	3
bone marrow																			
	normal	1	9	10	0	0	6	7	0	6	0	0	0/+	0	0/+	0	0	39	3
	anemic	1	16	13	0	0	12	15	0	6	0	0	0/NT	0	0/NT	0	0	63	10
thymus																			
	whole	1	2	2	2	15	0	0	0	0	3	1	0/+	2	0/+	0	0	28	6
	thymocyte (squeeze)	0	2	0	0	5	0	1	0	0	0	0	0/NT	0	0/NT	0	0	8	3
	stroma (squeeze)	2	0	1	0	2	1	0	0	0	0	0	0/NT	0	0/NT	0	0	6	2
	stroma (cyclophosph)	4	5	2	1	4	4	6	0	0	1	5	0/NT	0	0/NT	0	0	32	8
bursa																			
	whole	0	0	0	5	4	2	0	1	0	4	4	0/+	0	0/+	0	0	20	7
	stroma (cyclophosph)	0	0	0	6	6	0	0	0	0	5	10	0/NT	0	0/NT	0	0	27	5
intestine																			
	duodenum	0	0	0	11	3	0	1	9	1	11	11	0/+	6	0/+	0	0	53	2
	enterocytes (neonat)	0	0	0	0	0	2	0	0	0	8	0	0/NT	0	0/+	0	0	10	2
brain		3	0	2	0	17	1	0	0	0	0	0	0/-	0	0/+	0	0	23	6
kidney		0	0	1	17	0	0	0	0	0	2	0	0/-	5	0/+	0	0	25	2
liver		0	0	1	2	0	0	0	0	0	0	1	0/-	0	0/+	0	0	4	2
lung		3	4	0	1	0	3	3	0	0	0	0	0/-	0	0/+	0	0	14	2
heart		1	0	1	0	2	4	0	0	0	0	0	0/-	0	0/+	0	0	8	2
adrenal gland		2	4	4	6	2	2	1	0	1	0	0	0/-	0	0/+	0	0	22	1

Figure 1.9 The tissue-cell expression pattern of each BG gene of B12 haplotype (figure copied from Salomonsen *et al*., 2014).

The tissue or cell specific expression of each BG gene was determined by RT-PCR using primer UC74 and UC76, followed by cloning, sequencing and colony counting. The values in the table indicate the number of clones found by RT-PCR. The blue boxes represent haemopoietic BG genes (including BG5, 7, 8, 9, 11, 12 and 13) while the green boxes represent tissue BG genes (including the rest of the BG genes, BG1, 3, 4, 6 and 10) (Salomonsen *et al*., 2014).

1.2.7 BG functions

The functions of BG genes have been a mystery for half a century, although a number of perplexing immunological phenomena were observed. For example, the polymorphic BG determinants were responsible for the so-called 'natural antibodies' (found in unimmunized animals of many species); three other biological phenomena are described as following (Kaufman *et al*., 1991).

First, the BG antigens were observed with a so-called "adjuvant effect". A series of studies had showed that the injection of chicken B blood group antigens gave an obvious 'adjuvant effect' to the A blood group antigens when injecting partially inbred chickens with erythrocytes bearing different A or B antigens (Schierman *et al.*, 1967; Mcbride *et al.*, 1970, 1971). Later a study demonstrated that the 'adjuvant' was due to products of the BG region which promoted the response to BF antigens (Hala *et al*., 1981). A further experiment was performed using purified BF and BG proteins and proved that BG antigens were responsible for the 'adjuvant effect' (Salomonsen *et al.*, 1991). However, the mechanism is still unclear for how BG proteins function in such reactions.

Second, a BG protein found in intestinal brush border, originally known as the "zipper protein", might regulate the actin-myosin interaction using its cytoplasmic tail. Brush-border myosin 1 (BBM1) is a highly cell-specific form of myosin 1 molecule which binds to actin core and the membrane of the microvillus to fulfill the function of membrane transport; and tropomyosin regulates such activities by blocking the binding of BBM1 to actin. However, tropomyosin was not found in the microvilli where BBM1 was located, but instead, a 'zipper protein' was discovered by Bikle et al (Bikle *et al.*, 1993; Bikle *et al.*, 1996). Further study established that zipper protein is actually the cytoplasmic tail of BG protein, which regulates the binding of BBM1 to actin (Bikle *et al.*, 1996).

Third, BG1 is associated with resistance to certain viral diseases (Goto *et al.*, 2009). Two recombinant haplotypes derived from the same parental haplotype, BR2 and BR4, were observed with significant differences in resistance to Marek's disease virus (MDV) and Rous sarcoma virus (RSV) (Briles *et al*., 1980; Schat *et al*., 1994; Senseney *et al.*, 2000); Goto et al. mapped the crossover breakpoints that distinguish two haplotypes for gallid herpesvirus-2 (GaHV-2)-induced lymphomas (Marek's disease, MD) resistance, which pointed to the 3' UTR of the polymorphic BG1 alleles. A 225 bp insert in the 3' UTR of BG1 in BR4

haplotype was associated with increased lymphoma (Goto *et al.*, 2009). The authors also mentioned the possibility of the ITIM motif in the penultimate exon of BG1 determining resistance of Marek's disease virus (MDV) and Rous sarcoma virus (RSV). However, all these assumptions need to be tested, confirmed and understood in terms of the mechanisms.

1.2.8 Homologues

BG is a member of extended B7 family (Linsley *et al.*, 1994; Henry *et al.*, 1999; Salomonsen *et al.*, 2014). The extended B7 family concept was first proposed by Linsley et al. in 1994, which included the B7-related molecules, B7-1 and B7-2, butyrophilin (BTN), myelin oligodendrocyte glycoprotein (MOG) and chicken BG; and the common feature of such extended B7 family is the presence of at least one extracellular immunoglobulin (Ig) variable (V) domain, which shares significant amino acid sequence similarities (Linsley *et al.*, 1994). Subsequently, another two molecules, skin T cell (SKINT) and erythroid membrane-associated protein (ERMAP) were added into the extended B7 family (figure 1.10) (Henry *et al.*, 1999). More recently, Rhodes et al. defined chicken BG genes, together with most of the molecules mentioned above (ERMAP, MOG, SKINT) and all BTN and butyrophilin like (BTNL) molecules as the butyrophilin family (figure 1.10) (Rhodes *et al.*, 2016).

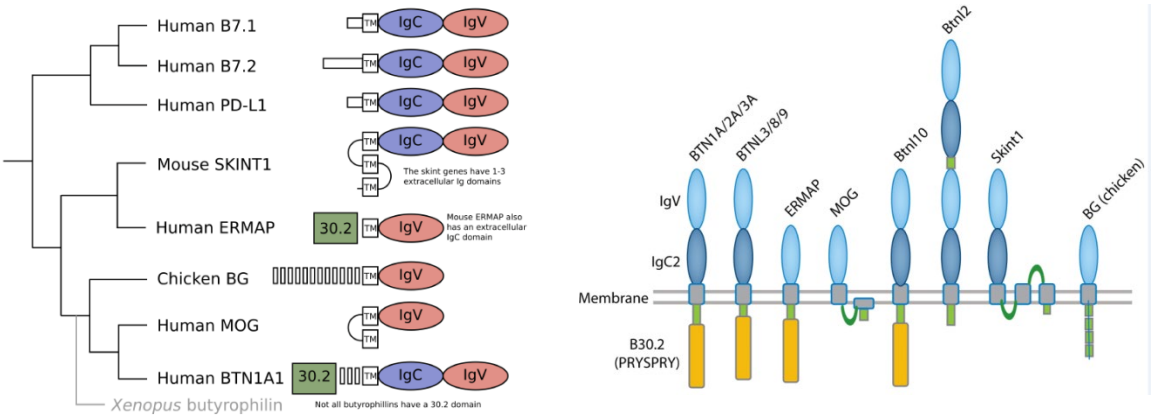


Figure 1.10 Members of extended B7 family and BTN family. The extended B7 family as portrayed originally by Henry *et al.*, 1999 and adapted by Chattaway, 2013 on the left, and the butyrophilin family created by Rhodes *et al.*, 2016 on the right. The phylogenetic tree on the left is built on amino acid sequence alignments of the full length proteins. Symbols are: IgC, immunoglobulin constant domain; IgV, immunoglobulin variable domain; TM, transmembrane domain; 30.2, 30.2 domain (Chattaway, 2013).

However, the chicken BG gene might be better classified in the extended B7 family rather than limited to butyrophilin family for two reasons. First, it is still unclear which is closer to the ancestor among BTN/BTNL, MOG and BG. The Ig-V domains of these three molecules share as high as about 40% similarities in amino acid sequences; therefore it is reasonable to believe they are from same ancestor. From the evolution perspective, a simple scenario would be that the common ancestor was the Ig-V domain, and during duplication, MOG gained transmembrane region, while BG and BTN/BTNL gained more domains, with cytoplasmic tail for BG and Ig-C and/or B30.2 domains for BTN/BTNL. Another possible scenario, MOG was the ancestor of BG and BTN/BTNL, and during duplication BG gained cytoplasmic tail, while some BTN/BTNL gained extra IgC and B30.2, some only gained extra IgC. Second, as shown in figure 1.2 above (section 1.1.2), chicken has the minimal essential MHC, while comparing the size and locations of BG genes in chicken genomes to BTN/BTNL in human, BG genes might have even more functions than BTN/BTNL.

The functions of these homologues, especially the BTN/BTNL families have been intensively studied recently, which might give some hints for BG functions. The BTN and BTNL molecules are emerging as novel regulators of immune responses in mouse and human (Dörner *et al.*, 2012). Some of them have been proved to co-inhibit T cell activation (Smith *et al.*, 2010; Dörner *et al.*, 2012; Ammann *et al.*, 2013; Ceeraz *et al.*, 2013; Rhodes *et al.*, 2015); and some of them are involved in the modulation of γδ T cells (Harly *et al.*, 2012; Vavassori *et al.*, 2013; Gu *et al.*, 2017). It has been demonstrated that the heteromeric pairing between BTNL6 and BTNL8, BTNL3 and BTNL8, and BTN3A1 and BTN3A2 occurs within the endoplasmic reticulum (ER), and is the conserved mechanism underpinning the selection and activation of γδ T cells in blood and extralymphoid sites in both mice and humans (Vantourout *et al.*, 2018). In addition, BTN genes have strong genetic associations with some diseases in human (Spagnolo *et al.*, 2007; Hsueh *et al.*, 2010; Hiramatsu *et al.*, 2011), and recent study shows several polymorphisms of butyrophilin family genes influence viral genotype selection in hepatitis C infection (Ampuero *et al.*, 2015). More recently, BTN3A was discovered being associated with tumors and the soluble isoforms of BTN3A1 was found in the plasma of pancreatic ductal adenocarcinoma (PDAC) patients (Benyamine *et al.*, 2018). The MOG is exclusively expressed in the central nervous system (CNS) (Jean, 2001) and contributes to the autoimmune-mediated demyelination (Bernard *et al.*, 1997). There are eleven mouse SKINT genes but in humans the single SKINT is a pseudogene (Boyden *et al.*, 2008). The mouse SKINT1 has been found to be a highly specific, uniquely selected

component for epidermal γδ T cells (Boyden *et al.*, 2008; Barbee *et al.*, 2011). Currently, there are no reports identifying a chicken SKINT gene.

1.3 Aims of this project

In order to understand the evolutionary history and potential functions of chicken BG genes, there are still a lot of works to do. First of all, it is important to understand the BG genes in chicken haplotypes other than B12, especially the BG genes in the BG regions. So far, only BG genes in B12 haplotype have been fully typed and analyzed; it is very time consuming and costly to repeat the same work for other haplotype. PCR would be a good option to explore BG genes in other haplotypes; however, the difficulty is the design of PCR primers due to the sequence polymorphism. After many efforts, there was only one primer pair called 'H2' available, which was designed based on the BG genes from B12 haplotype that have the haemopoietic 5' UTR and type 2 3' UTR. H2 primer pair worked well on different tissues from five haplotypes (Chattaway, 2013). Second, it is important to identify which BG genes in the BG region from different haplotypes are alleles. Due to the polymorphism and CNV, there is no idea about orthologous BG genes in the BG region. However, such information is critical to understand the evolution as well as the function. For example, by comparing the allelic BG gene sequences, it will be seen which region of the gene is conserved, whether the polymorphism is selected, and if so, what is it selected for, etc. Third, the real BG transcripts should be systematically studied. Alternative splicing has been mentioned by several groups from different angles, but systematic study of BG cDNA sequences is lacking, which is critical for understanding the BG biology. Last but not the least, reagents including BG mAbs, BG-Fc fusion proteins etc., are essential for further characterizing BG proteins and exploring BG functions; therefore, it is urgent to develop these reagents.

To solve as many problems addressed above as possible, this project includes three subjects with specific aims as following.

Subject 1, to develop a PCR method which could amplify nearly full-length cDNA sequences of all haemopoietic BG genes from different chicken haplotypes. It is unrealistic to have a pair of universal primer amplifying all haemopoietic and tissue BG genes due to the high polymorphism. Since most studies and observations about BG antigens are from erythrocytes, the first step of understanding the BG gene expression at the cDNA level would best start with haemopoietic BG genes. H2 primers, aiming to amplify all haemopoietic BG genes with

type 2 3' UTRs, have been tested on some tissue samples; therefore, at least another primer pair 'H1', amplifying all haemopoietic BG genes with type 1 3' UTRs, should be developed. Ideally, if a universal primer pair for all haemopoietic BG genes could be designed, it would represent progress for exploring BG genes.

Subject 2, to identify the 'functional alleles' of BG genes in T cells and B cells of different chicken haplotypes. As discussed previously, according to the tissue distribution work on B12 haplotype, specific BG genes were dominantly expressed in certain cell types. For example, in B12 T cells there is BG7 dominantly expressed with a few clones of BG12; in B12 B cells there are two virtually equally expressed BG genes, BG7 and BG9, with several sub-dominantly expressed genes, BG8, BG12 and BG13. These dominantly expressed BG genes in particular cell type from different haplotypes should have the same functions, so they could be treated as 'functional alleles'. Once the 'functional alleles' were identified, sequences could be compared to answer those questions addressed in the beginning. In addition, T cells and B cells are among the most important cell types in immune system. To understand which BG genes and which transcripts expressed in T and B cells will help studying functions of BG genes. Also, based on B12 haplotype, almost all of the haemopoietic BG genes (five out of six) were found in T and B cells; therefore, hopefully through this project we could clone most of the haemopoietic BG genes in other haplotypes, which would contribute a lot of sequence information for further study using these haplotypes.

Subject 3, to develop reagents for further characterizing BG proteins and exploring BG functions. BG-Fc fusion proteins will be developed first, then they will be used to screen BG monoclonal antibody (mAb) supernatants made by Prof. Jan Salomonsen. Once BG mAbs are characterized by enzyme-linked immunosorbent assay (ELISA) and western blot, specific BG mAbs can be used for tissue distribution study and further function study. In the meantime, BG-Fc fusion protein could also be applied to functional assays and ligand/receptor searching.

Chapter 2

General materials and methods

2.1 Chicken lines, samples and cells

2.1.1 Chicken lines and haplotypes

Four genetic lines of White Leghorn chickens were maintained under specific pathogen-free (SPF) condition at the Institute for Animal Health (IAH) Compton UK (now rebranded as the Pirbright Institute): line 6_1, 15I, N and P2a. Among these, the congenic line 6_1 and 15I were inbred lines developed and maintained since 1939 at the Regional Poultry Research Laboratory (RPRL) at East Lansing, MI USA, with the original names of RPRL-6_1 and RPRL-15I respectively (Stone *et.al*., 1975). The congenic line N was originally developed by Cole at Cornell University, Ithaca, NY USA and introduced to the Avian Disease and Oncology Laboratory (ADOL), MI USA, then to IAH (Bacon *et.al*., 2000). The genetic line P2a was also originally developed at Cornell University (I Shaw *et al.*, 2007) and introduced to the Institute for Animal Science and Health, Lelystad, the Netherlands, then to IAH (Bacon *et.al*., 2000).

These chicken lines above were originally developed for resistance or susceptibility to particular viral diseases (Table 2.1) (Stone *et.al*., 1975; Bacon *et al*., 2000). They were later typed for their major histocompatibility (B) complex specificity (BF, BL and BG loci) and acquired a formal nomenclature (Table 2.1) (Briles *et al.*, 1982; Miller *et al*., 2004).

Table 2.1 Chicken lines, haplotypes and disease resistance

Chicken line	MHC B haplotype	BL	BF	BG	Disease resistant or susceptible			
					MDV	LLV	ALV	RSV
6_1	B2	2	2	2	R	R	S	S
15I	B15	15	15	15	S	S		
P2a	B19	12=19	19	19	S			
N	B21	21	21	21	R			

MDV: Marek's disease virus; LLV: lymphoid leukosis virus; ALV: avian leukosis virus; RSV: rous sarcoma virus. Disease resistance summarized from Stone *et al*., 1975 and Bacon *et al*., 2000.

2.1.2 Samples

The spleen, ileum and peripheral blood samples were collected after schedule 1 killing of single animals by Dr. Colin Butter and Ms Karen Staines at the Pirbright Institute. The duodenum sample used in this experiment was obtained from a SPF chicken line 15I and stored in -80°C with RNAlater (Sigma) (kind gift from Dr. Colin Butter).

2.1.3 Cells

Human embryonic kidney cells 293 (HEK293) and HEK293 derived stable cell lines were maintained in Dulbeco's modified Eagle's medium (DMEM) with 10% fetal bovine serum (FBS) and antibiotics (Appendix A1). All cell lines were split once a week at a ratio of 1:4 by a standard trypsinisation method. Briefly, the medium from monolayer cultures was decanted and the monolayer rinsed with PBS (Appendix A1) and the culture flask emptied. A pre-warmed trypsin solution (Appendix A1) was added and poured off, leaving sufficient trypsin solution to just cover the monolayer. The culture flasks were then incubated at 37°C for three to five minutes and then the cells were dispersed by banging of the flask and pipetting. The cells were then resuspended in an appropriate volume of cell culture medium and dispensed into new flasks. The cultures were maintained in a controlled environment incubator at 37°C with an atmosphere of 5% CO_2 and >90% humidity.

2.2 Methods and protocols

2.2.1 RNA extraction

2.2.1.1 Tissue RNA extraction

The general principles of RNA extraction were followed from the manufacturer's protocol (NucleoSpin® RNA II RNA extraction kit, Machery-Nagel), while extra protocols were implemented in order to get high quality RNAs. Before handling tissues, the area where the RNA extraction procedure would be performed, and the pipettors were treated with RNase removing agent (RNAseZAP®, Sigma). About 25 mg tissues were homogenized using 1.5 mL sterile pellet pestle (Sigma), after which 350 µl lysis buffer was added immediately to avoid RNA degradation from potential RNase existing in the tissues. The lysate was

transferred into the blue column provided with the kit and centrifuged to filtrate proteins. The lysate was adjusted by adding 350 µl 70% ethanol and transferred into the RNA binding column for centrifugation. The RNA binding column was desalted using 350 µl membrane desalting buffer (MDB) to ensure DNAs can be most effectively removed in the following DNA digestion step which the column was treated with DNase reaction mixture [90 µl reaction buffer with 10 µl rDNase (a recombinant version of bovine DNase) solution]. The column was then washed three times with the sequence of 200 µl buffer RAW2, 600 µl buffer RA3 and 250 µl RA3, and finally the RNAs were eluted with 50 µl RNase-free water and stored at -80°C until being used.

2.2.1.2 Cell culture RNA extraction

When extracting the RNAs from sorted cells or cell cultures, cell numbers were counted before applying the extraction procedure. Roughly $1x10^6$ cells were enough to provide sufficient RNAs. Cells were washed quickly using PBS, centrifuged to discard the liquid and then resuspended with lysis buffer immediately. The rest of the procedures were the same as the '2.2.1.1 Tissue RNA extraction' above.

2.2.2 First strand cDNA synthesis

In a previous comparison of cDNA synthesis kits, the 'Maxima H Minus First Strand cDNA Synthesis Kit (ThermoFisher)' had been found superior in creating full length BG cDNA (Chattaway, 2013). It appears that BG RNAs have robust secondary structure which requires hotter temperature to break down in order to convert mRNAs to cDNAs, while the modified M-MuLV reverse transcriptase provided in this specific kit above is capable to synthesize cDNA at higher temperature (up to 65°C) without RNA degradation compared to other kits being performed at 42°C. Therefore, the cDNA synthesis protocol was inherited mostly from Dr. John Chattaway who primarily optimized the incubation temperature (55°C) with strictly following the manufacture instruction, while I extended the incubation time from 30 min to 45 min.

To summarize the whole procedure, RNAs were first mixed with oligo $(dT)_{18}$ primer and dNTP mixtures, heated at 65°C for 5 min, then chilled on ice for 3 min before adding RT buffer and Maxima H Minus Enzyme Mix to complete reaction mixtures. The reaction mixtures were then incubated at 55°C for 45 min to generate first strand cDNAs, followed by

85°C for 45 min to inactivate the enzyme which might affect following PCR. For longer storage (more than two months), the cDNAs were kept at -80°C or -20°C before use.

2.2.3 Primer design

For normal PCR reactions, the primer designing programme 'Primer3Plus' (http://primer3plus.com/) was utilized to design most of the primers and to evaluate GC content, secondary structure and melting temperature (Tm). When primers were designed manually, three key considerations were followed to ensure successful design: first, the primer length was kept around 20 nts without counting the restriction enzyme site in the 5' end; second, the last 6 nucleotides of the 3' end were highly conserved to the template; third, the last 3 nucleotides of the 3' end were either C or G in order to promote the binding affinity.

When primers were designed for cloning purpose, apart from the considerations in normal PCR primer design above, extra bases were added to the 5' end of the restriction enzyme site to protect the cleavage site. Recommendations regarding which bases should be added to particular restriction enzyme sites were listed in NEB website (https://www.neb.com/~/media/NebUs/Files/Chart%20image/cleavage_olignucleotides_old.pdf).

To design the primers for introducing mutations, an online automatic primer design tool called NEBaseChanger was used (http://nebasechanger.neb.com/). By entering the sequence and naming the mutation site, the primer candidates would be listed.

Primers used in this thesis were manufactured by Sigma-Aldrich, stored in TE buffer with the concentration of 100 μM, and were diluted into working solution (10 μM) using nuclease free water whenever being used. The oligo information of all the primers used in this thesis is listed in appendix B.

2.2.4 Polymerase chain reaction (PCR)

Different PCR kits and procedures were used appropriate to various experimental purposes; however, the general rules for PCR amplification were the same. The formulae to set up a PCR reaction require dNTPs, buffer solution, one primer pair, DNA template, and DNA polymerase which was either standard thermostable DNA polymerase called Taq polymerase or a high-fidelity polymerase called Phusion polymerase that has a 3'-5' exonuclease activity

and can remove erroneously incorporated bases. The procedure of PCR amplification was performed on BIO RAD PCR machine (Alpha™ Unit Block Assembly for DNA Engine® System) and contained the following steps: the initialization step, to allow the double-stranded DNA to be separated into two strands, the cycling steps (normally 25-35 repeats of a cycle including denaturation, annealing and extension), and the final elongation step, to ensure all the DNAs are extended fully.

PCR is now a conventional lab technique; however, some tips are essential for a successful reaction, especially for these experiments where consistency is required. For this thesis, the following four protocols were consistently followed. First, all the reagents were thawed totally and inverted twice, to ensure they were mixed well. Second, a master mixture without templates was prepared, mixed well and aliquoted into individual PCR tubes when large numbers of reactions required. Third, the reactions were set on ice. The last, both negative and positive controls were included as part of each experiment.

2.2.5 DNA gel electrophoresis

PCR products were analyzed and separated by DNA gel electrophoresis. The gels used in this thesis were made by dissolving 0.8-1% agarose (UltraPure™, Invitrogen) into 1x Tris acetate (TAE) buffer by microwave (at high temperature for 5 mins); then the solution was cooled down, mixed with Sybr Safe (Invitrogen) to a final concentration of 0.5 μg/mL and poured into the appropriate casting tray with suitable comb inserted into the cast tray slots. The combs were carefully pulled out when the gels were solidified, and the gels were then placed into the electrophoresis tank filled with 1 x TAE buffer. The PCR products mixed with 6 x loading buffer (NEB) were loaded into wells, so were the 1 kb and 100 bp DNA ladders (GeneDireX®). The gels were run under 110 V for 30-50 mins according to different experimental purposes and the DNA fragments were checked by visualizing the gels under UV light or FluorChem 9000 gel imager (Alpha Innotech).

2.2.6 PCR products purification and DNAs isolation from agarose gels

ISOLATE II PCR and Gel Kit (Bioline) was used both for purifying PCR products and for isolating DNA fragments from agarose gels. For the former, two volumes of binding buffer CB were added to the PCR tube and were mixed well with one volume of PCR product. For the latter, at least 200 μl binding buffer CB was added to dissolve 100 mg of agarose gel. The

whole mixture was then transferred into silica membrane column provided in the kit, centrifuged for 1 min at 11,000 x g to remove the waste and to allow the DNAs to bind to the silica membrane. After washing using 700 µl washing buffer CW and centrifugation at 11,000 x g for 30 s, the silica membrane was dried and the DNAs were eluted with 15-80 µl elution buffer TE by centrifuging at 11,000 x g for 1 min. DNAs were kept at -20°C for short time storage, and at -80°C for longer storage.

2.2.7 DNA precipitation

The ethanol precipitation method was used to concentrate DNAs and all the reagents were chilled on ice before use. A 1/10 volume of sodium acetate (3 M, pH 5.2), 1/10 volume of $MgCl_2$ (0.33 M) and 2 volumes of ethanol (at least 95%) were added to the DNA aqueous solution, incubated on ice for 30 mins followed by centrifugation at 17,000 g, 4°C for 10 mins. The supernatant was removed gently using a pipette, and the DNA pellet was rinsed by adding 400 µl ethanol (at least 95%) and spun again at 17,000 x g, 4°C for another 10 mins, after which the supernatant was discarded using a pipette and the DNA pellet was dried at 50°C for 3 mins and dissolved in the desired buffer.

2.2.8 Site-directed mutagenesis

Site-directed mutagenesis was applied to substitute or delete a nucleotide and also to insert a large DNA fragment into a plasmid; and the Q5® Site-Directed Mutagenesis Kit (NEB) was used to perform the reaction following the manufacturer's instructions. The whole procedure could be divided into four main steps: primer design, PCR reaction, Kinase-Ligase-DpnI (KLD) treatment and transformation.

For primer design, the online tool NEBaseChanger (http://nebasechanger.neb.com) was used to help in designing primers, which was simply achieved by inputting the plasmid sequence, choosing the programme (whether for substitution, insertion or deletion), annotating the exact position in the sequence where the mutation would be taken and inputting a DNA sequence if it was for insertion. The tool automatically recommends primer pairs with primer property information, e.g. Tm, GC content etc., which was useful for setting the PCR reaction programme later in the PCR amplification step. Primer oligonucleotides were synthesized and purified using the standard method by Sigma, with no special requirement, unlike many other mutagenesis kits.

For the PCR amplification, a total volume of 25 µl reaction mixture was prepared by mixing 9 µl nuclease-free water, 1.25 µl forward primer (final concentration 0.5 µM), 1.25 µl reverse primer (final concentration 0.5 µM), 12.5 µl 2x Master Mix and 20 ng template. The reaction mixture was spun and placed on BIO RAD PCR thermocycler following the PCR cycling programme of 30 seconds of initial denaturation at 98°C, 25 cycles of denaturation (98°C 30 seconds), annealing (at a particular temperature, depending on the primer pair, for 20 seconds), and extension (72°C 20 seconds/kb), and 10 mins of final extension at 72°C.

The PCR product aliquots (1 µl) were then treated with KLD reagents by mixing with 3 µl nuclease-free water, 5 µl 2x KLD reaction buffer and 1 µl 10x KLD enzyme mix, and incubated at room temperature for 5 mins. Then, the PCR products were transformed into competent cells (*E. coli* DH5α) using the exact same protocol as described in 2.2.9.

2.2.9 DNA ligation and transformation

Both sticky-end ligation and blunt-end ligation techniques were performed in this thesis. Generally, the sticky-end ligation was applied when DNA fragments needed to be inserted into a particular plasmid where double digestion had been done on both the DNA fragments and vector plasmid using specific restriction enzymes; while the blunt-end ligation was mainly used for the purpose of sequencing.

T4 DNA Ligase kit (NEB) was used for sticky-end ligation by adding 2 µl of 10x reaction buffer, 1 µl of T4 DNA ligase (final concentration 10 U/ µl), 50 ng of vector, 3x molar DNA fragments compared to the vector amount, and adjusting the volume using nuclease free water to make a total reaction volume of 20 µl. The reaction was carried out at room temperature for 10 mins.

CloneJET PCR Cloning Kit (Thermo Scientific) was used for blunt-end ligation. The reaction was prepared according to the manufacturer's instructions by adding 10 µl of 2x reaction buffer, 1 µl of T4 DNA ligase (final concentration 5 U/ µl), 50 ng of vector, 150ng DNA fragment, and adjusting the total volume to 20 µl using nuclease free water.

The whole fresh ligation products were transferred into 50 µl aliquots of completely thawed competent cells *E. coli* DH5α cells, avoiding vigorous mixing but gently flicking the tube. The competent cells were then chilled on ice for 30 mins, heat shocked at 42°C for 30

seconds, chilled on ice for another 5 mins, and recovered by growing with 950 µl LB cultures in a shaking incubator agitating at 250 rpm at 37°C for 1 hour, after which 150 µl cultures were aliquoted and spread onto pre-warmed homemade LB-ampicillin agar plates to grow colonies overnight in incubator at 37°C. Colonies were picked by gently touching the isolated clones using sterile 10 µl tips, transferred to 50 µl LB cultures and incubated for two to three hours at 37°C. Then 1 µl aliquot of each LB culture was used to run colony PCR and only the clones with positive results in colony PCR were grown up in LB-ampicillin cultures.

2.2.10 Colony PCR

Colony PCR was used in the work described to confirm the successful insertion of DNA fragments into plasmid vector before sequencing. Compared to conventional PCR, there were three major considerations: preparation of template, polymerase and primers. First, in colony PCR, the colonies were picked by gently touching the isolated clones using sterile 10 µl tips and transferred into individual 0.5 mL micro tubes containing 50 µl LB-Ampicillin cultures. The cultures were then incubated at 37°C for 2 to 3 hours, and 1.5 µl aliquot cultures were used as PCR template. Second, the Taq DNA Polymerase Recombinant kit (Invitrogen™, ThermoFisher) was used. Because the Taq DNA polymerase is heat resistant, it is still active after the initial denaturing step (94°C 5 mins), and therefore facilitated amplification of the plasmid DNA. Third, either primers targeting the inserted DNA fragment or primers targeting on the vector upstream and downstream of the inserted DNA fragment could be used for colony PCR. However, in some cases, for example the DNA fragment inserted had not been confirmed by sequencing, primers targeting on DNA fragment would help to reduce possibility of choosing false positive colonies.

The same recipe and PCR programme were applied in all colony PCR in this thesis. A total of 25 µl reaction reagents were made by adding 19.65 µl nuclease free water, 0.5 µl 10 mM dNTP mix (2.5 mM each dATP, dTTP, dCTP, dGTP), 0.75 µl 50 mM $MgCl_2$, 0.5 µl 10 µM forward primer, 0.5 µl 10 µM reverse primer, 2.5 µl 10x PCR buffer, 0.1 µl 5 U/µl Taq DNA Polymerase and 1.5 µl colony cultures, vortexed for 3 seconds and spun down, then placed on PCR machine to be denatured at 94°C for 5 mins, then followed by 30 cycles of PCR circle which including denaturation at 94°C for 45 s, annealing at 56°C for 30 s, and extension at 72 °C for 90 s, and final extension at 72 °C for 10 mins.

2.2.11 Miniprep

Plasmids were purified from bacteria *E. coli* DH5α using PureLink® Quick Plasmid Miniprep Kits (Invitrogen). Bacteria were cultured overnight in LB-Ampicillin cultures and harvested by centrifugation at 1500x g for 10 mins and removal of the supernatant medium. The bacteria were then resuspended using 250 µl R3 buffer provided by the kit, transferred into 2 mL sterile Eppendorf tube, and lysed using 250 µl lysis buffer (L7) by inverting the tube 5-8 times for mixing following by incubation at room temperature for 5 mins. The lysed bacteria were then neutralized by adding 350 µl buffer N4 and immediately shaken to make the mixture homogeneous. The mixture was then centrifuged at 17,000x g for 15 mins, and the clean supernatants, generally around 800 µl, were transferred into silica column provided by the kit, and centrifuged at 11,000x g for 1 min, allowing the plasmid to bind to the membrane of the silica column. The silica column was washed using 700 µl washing buffer CW9 and centrifuged at 11,000x g for 1 min, followed by a second centrifugation with the column placed into a new collecting tube to make sure all the washing buffer was totally discarded. The silica column was placed on a new sterile 1.5 mL micro tube and treated with 15-100 µl of the desired elution buffer, according to experimental purpose, to elute the plasmids. Finally, the plasmid was measured using a Nano Drop ND-1000 (Nanno Drop Technologies) and stored at -20°C for short time and -80°C for long term.

2.2.12 Sequencing and data analyzing software

The sequencing samples, primarily plasmids, were prepared either in 2 mL centrifuge tubes (10 µl each sample in each tube) or in a 96 well PCR plate for high throughout samples (2 µl each sample in each well) with the concentration around 70-120 ng/µl. All the samples were sent to the DNA sequencing facility in the Department of Biochemistry, University of Cambridge for sequencing using the Sanger sequencing method.

The sequencing data was viewed in CLC DNA Workbench 5 (CLC), and for sequence assembly, SeqMan (DNASTAR®) was used. Consensus sequences generated by SeqMan were exported as a single sequence and imported to CLC for other analyses.

2.2.13 Transfection

HEK293 cells were transfected with various plasmids according to different experimental

purposes but all procedures used the jetPEI® DNA Transfection Reagent (Polyplus Transfection) following the manufacturer's protocol. The jetPEI® DNA Transfection Reagent is a chemical-based transfection technique, where the plasmids can be compacted by the jetPEI, a polythylenimine, into positively charged particles that interact with anionic proteoglycans at the cell surface and enter cells by endocytosis (Akinc *et al*., 2005). The optimized protocol was applied into each transfection as follows.

HEK293 cells were split equally into 6-well plate after being passaged 2-3 times after being thawed, and were ready to receive the plasmids/jetPEI complex when the cells were 80% confluent. The plasmids/jetPEI mixture (for one well of 6-well plate) was made by mixing plasmids solution (2 ng plasmid diluted into 50 µl 150 mM NaCl) and jetPEI solution (10 µl jetPEI diluted into 50 µl 150 mM NaCl), and incubated at room temperature for 25 mins. Then the plasmids/jetPEI mixture was added to the cells with 2 mL fresh medium and mixed with the cells by gently swirling the plates. The transfected cells were then incubated in 37°C for 32 hours. In the transient transfection experiment, cells or supernatants were suitable for further experiments 32 hours after being transfected. For experiments employing stable transfection, the cells were transferred into 25 cm^2 flask containing 5 mL fresh medium and 100 µl 50 mg/mL Geneticin 418 (G418) (Santa Cruz Biotechnology) 32 hours post transfection for selection. During the selection period, most cells would die, and fresh medium with G418 was changed gently every 5-7 days until cells were around 30-50% confluent, after which cells were ready to be split and passaged also using medium with G418. Cells were checked for protein expression at each passage by Western blot, to make sure that the inserted gene was well expressed. If the inserted gene was expressed stably for at least three passages, the cells could be considered as stable cell lines, thus were frozen in vials with freezing medium (90% FBS and 10% DSMO) overnight at -80°C and transferred to liquid nitrogen tank. Otherwise, transfection was repeated until the inserted gene was stably expressed as described above.

2.2.14 SDS-PAGE and Western blot

Formulae for reagents used for sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) are given in appendix A3. Protein samples were mixed with same amount of 2x SDS loading buffer for non-reduced SDS-PAGE, and together with DTT (to a final concentration of 200 mM) for reduced SDS-PAGE, heated at 95°C for 5 mins and, once cooled, were loaded into the SDS-PAGE gel (4% stacking and 10% separating gel). The SDS-

PAGE gel was prepared in advance following the standard protocol (Appendix A3), placed into the Bio Rad apparatus tank with 1x SDS running buffer and, after being loaded with the protein marker, a mixture of XP Western Blot Standards (Invitrogen) and Kaleidescope Prestained Protein Standards (BioRad) as well as the protein samples, the SDS-PAGE gel was run at 200 V for 50 mins. The SDS-PAGE gel was then trimmed, quickly rinsed with Transferring Buffer and the proteins in the SDS-PAGE gel was semi-dry transferred into Immobilon-P Membrane (Merck Millipore) using Trans-Blot SD Semi-Dry Transfer Cell (Bio Rad) at 120 V/gel for 47 mins.

The membrane transferred with proteins was blocked at room temperature for 1h or at 4°C overnight using 1x Tris buffer saline with tween 20 (TBST) with 1% bovine serum albumin [BSA, Sigma, ≥ 96% (agarose gel electrophoresis)]. After being washed three times with 1x TBST, the membrane was incubated with primary antibody diluted with the recommended ratio according to the manufacture protocol in 1x TBST with 1% BSA at room temperature for 1 h or at 4°C overnight. The diluted antibody was collected and frozen at -20°C for future use and the membrane was rinsed for 4 x 8 mins each with 1x TBST, which was then incubated with the secondary antibody containing HRP diluted with the recommended ratio according to the manufacturer's protocol in 1x TBST with 1% BSA at room temperature for 1 h or at 4°C overnight. After being washed for 4 x 8 mins each with 1x TBST, the membrane was stained with chemiluminescent detection reagent ECL (GE Healthcare, Life technology) and exposed using X-ray Film (Konica Minolta).

2.2.15 Cell sorting

Approximately $5x10^7$ lymphocytic cells isolated from peripheral blood or particular tissues were counted, transferred into a 15 mL centrifuge tube which was filled with cold PBS (chilled on ice) to make the total volume of 10 mL, and centrifuged at 4°C for 5 minutes (300 g) and the supernatant was removed leaving the cell pellet in the bottom. The cell pellet was washed again by resuspending with 10mL cold PBS, centrifuging at 4°C for 5 minutes (300 g) and discarding the supernatant. Then the cell pellet was resuspended with 2 mL cold PBS, of which 0.5mL of each cell suspension was collected as pre-sorting control (or negative control) and the rest of each cell suspension was stained with particular PE conjugated antibodies as described below.

In this thesis, three different cell types were stained for sorting, which were T cells, B cells

and intestinal epithelial cells (IECs). To sort these cells, cell suspensions were first stained with specific R-phycoerythrin (RPE) conjugated antibodies: 10 µL mouse anti-chicken CD4-RPE (SouthernBiotech) and 10 µL of mouse anti-chicken CD8-RPE (SouthernBiotech) for T cell sorting, 10 µL mouse anti-chicken Bu-1a-RPE (SouthernBiotech) for B cell sorting, and 10 µL mouse anti-chicken CD45-RPE (SouthernBiotech) for IEC sorting. Then the cells were incubated at 4°C in the dark for 1 h, after which they were washed three times by adding 10 mL cold PBS, spinning at 4°C for 5 minutes (300 g) and discarding the supernatant. Finally, the washed cells were resuspended into 1 mL cold PBS for sorting.

The sorting procedure was operated by flow cytometrist Mr. Nigel Miller in the Department of Pathology using a DakoCytomation MoFlo MLS high-speed cell sorter (Beckman Coulter). T and B cells were sorted as positive outcomes because the antibodies used recognize cell surface molecules of T and B cells, while IEC were sorted as negative outcomes since epithelial cells do not express CD45 molecules.

Chapter 3

A novel PCR protocol to amplify BG transcripts from haemopoietic cells

3.1 Introduction

Though having been discovered for half a century, there remains much to be understood about BG: what does BG do, why BG genes are so polymorphic etc. Lots of interesting suggestions and hypotheses regarding BG function and evolution in immune system were proposed by researchers; however, none of them has been tested so far. To be able to answer any of the questions above, one fundamental problem needs to be solved urgently to help us go further, which is to understand the expression of BG genes.

Previous methods used to explore the full length BG cDNA transcripts were mainly focused on building cDNA libraries. For example, Prof. Jim Kaufman found eight clones from a cDNA library made from anemic B19 bone marrow showing BG genes although they all only contained partial cDNA sequence (Kaufman *et al*., 1989). Later five nearly full-length BG cDNA were identified through constructing cDNA library using erythroid cells of a B21 haplotype chicken (Miller *et al.*, 1991). There is no doubt that cDNA libraries have provided useful and essential resources for getting cDNA transcripts; however, constructing a cDNA library is difficult, tedious and quite often the alternatively spliced isoforms cannot be well explored compared to the dominant transcripts (Harbers, 2008). Later, both Prof. Kaufman and Prof. Miller's lab developed their own PCR protocols to amplify BG cDNA sequences and submitted many BG sequences to GenBank, but none of them were full length cDNA sequence (Salomonsen *et al*., 2014; Iglesias *et al*., 2007).

With the development of genomic sequencing, various methods were used to identify BG genes and a significant milestone was the annotation of all 14 BG genes from B12 haplotype chicken (Salomonsen *et al*. 2014). In the meantime, the two singleton BG genes (BG0 and BG1) were studied in both Prof. Kaufman and Prof. Miller's labs. In Prof. Kaufman's lab, specific primer pairs to amplify BG0 and BG1 genes were designed, and both BG0 and BG1 alleles from different chicken haplotypes were examined (Chattaway *et al.*, 2016). In Prof. Miller's lab, the chicken MHC regions of different haplotypes were typed, from which the genomic sequences of BG1 genes were identified (Hosomichi *et al.*, 2006, 2008).

After gaining much BG sequence information, PCR would be the obvious tool to examine cDNA sequences. However, due to the high polymorphism and copy number variations of BG genes in BG region, the attempt of designing a universal primer pair amplifying all BG genes failed (Chattaway, 2013). So far, the best solution proposed by Dr. John Chattaway was to develop four primer pairs (H1, H2, T1 and T2) described as follows.

As mentioned previously in section 1.2.6 chapter 1, the phylogenetic tree built on 5' UTRs of all 14 B12 BG genes showed that all the genes fell into two groups which correlated to tissue distribution: genes in one group were found in haemopoietic cells thus called Haemopoietic BGs, while genes in the other group were found in tissues thus called Tissue BGs. Although BG genes are highly polymorphic, the 5' UTR sequences of the genes from the same group were comparatively conserved. Therefore, in theory, one universal forward primer could be designed to amplify Haemopoietic BGs and another one for Tissue BGs. Also, the phylogenetic tree built on 3' UTRs showed two different clusters as well, so called type 1 BGs having type 1 3' UTRs and type 2 BGs having type 2 3' UTRs, which should allow a universal reverse primer to be designed for type 1 BGs and another one for type 2 BGs. Thus, the idea of designing four types of primer pair (H1, H2, T1 and T2) to amplify all BG genes was formed, with the letter H or T for BG genes expressed by haemopoietic cells or expressed in tissues respectively, and the number 1 and 2 for BG genes whose 3' UTR are type 1 or type 2. To ensure the four primer pair idea work for as many chicken haplotypes as possible, almost every 5' UTR and 3' UTR sequences available were aligned, and Dr. Chattaway started the project by designing one primer pair (H2) first.

H2 primers were designed to amplify Haemopoietic BG genes with type 2 3' UTRs. As shown in figure 3.1, all the Haemopoietic 5' UTRs and type 2 3' UTRs could be targeted by the H2 primers (forward primer ID UC206 and reverse primer ID UC210) from the sequence alignments. Later, a PCR protocol based on H2 primers (H2-PCR) was established by Dr. Chattaway, and successfully applied to investigate BG genes from four tissue samples (spleen, thymus, bursa and duodenum) with each from five different chicken haplotypes (B2, B4, B12,

B15 and B21), from which 20 new BG genes were found (Chattaway, 2013).

One part of my project is to systematically examine the BG cDNA sequences in T and B cells of different chicken haplotypes; as T and B cells are haemopoietic cells, I could amplify all BG genes from the two cell types by using H2-PCR and develop H1-PCR. In order to get myself familiar with H2-PCR protocol, I started my project by helping Dr. Chattaway finish his work: investigating BG genes from duodenum sample of B15 haplotype chicken using H2-PCR protocol. Through this work, the H2-PCR method was re-evaluated by myself after facing many difficulties in practice, and a new PCR protocol (HU-PCR) which can amplify all Haemopoietic BG genes with both type 1 and type 2 3' UTRs was developed after overcoming these difficulties.

In this chapter, the work of exploring new BG genes from duodenum sample of B15 haplotype using H2-PCR protocol was described first. The problems I met through this practice, how the problems were solved and the results were then described. An improved protocol, called HU-PCR, was explained, including what the new protocol was, how it worked, why it performed better compared to H2-PCR and why it was chosen to replace H2-PCR for examining haemopoietic BG genes.

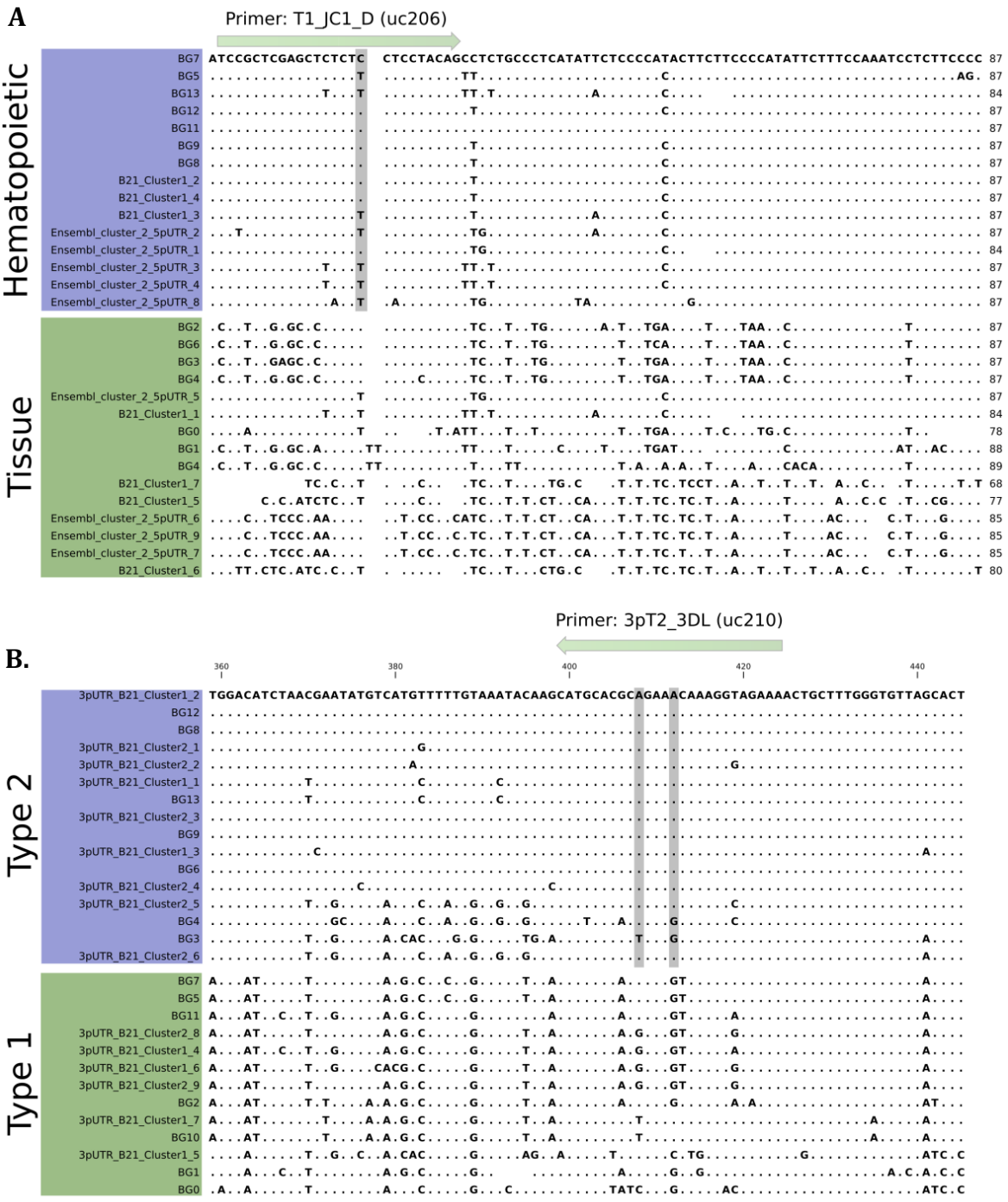


Figure 3.1 Locations of H2 primers on the nucleotide sequence alignments of BG genes from B12 and B21 haplotypes (figure adapted from Chattway, 2013). A. The alignment of 5' UTRs from the B12 and B21 BG genes. The forward primer of H2 (UC206: 5' TCCGCTCGAGCTCTCTYCTCCTACAG 3') is annotated as an arrow and the degenerate base which is highlighted in the vertical grey bar was Y (for C and T). B. The alignment of 3' UTR from the B12 and B21 BG genes. The reverse primer of H2 (UC210: 5'TTTCTACCTTTGYTTCSGCGTGCATG 3') is indicated as an arrow and the degenerate bases which are highlighted in the vertical grey bars were R (for A and G) and W (for A and T) (Chattaway, 2013).

3.2 Materials and methods

3.2.1 Chicken lines, haplotypes and samples

Three different chickens lines were used to establish and optimize the standard protocol for cDNA transcripts obtained for BG genes, which were line 15I, line P2a and line N with the MHC haplotypes of B15, B19 and B21, respectively. The sample used for RNA generation was duodenum tissue from line 15I, which was stored in -80°C with RNAlater (Sigma) (kindly gift from Dr. Colin Butter); and sorted intestine epithelial cells (IECs) from line N and line P2a. The IECs were separated from ileum samples by Dr. Michaela Fakiola and myself following the protocol in the following section 3.2.2, and sorted by flow cytometrist Mr. Nigel Miller in Department of Pathology. The RNA generated from sorted peripheral T cells and B cells of line N and line P2a after MACS cell sorting (Miltenyi Biotec) was given by our collaborator Ms Karen Staines and was confirmed with the RNA quality by PCR amplification of housekeeping gene GAPDH.

3.2.2 IECs separation protocol

The IECs separation protocol was adapted from the original one from Dr. Marc Veldhoen in Babraham Institute, Cambridge UK. The ileum sample (the last two fifths in the distal part of small intestine) was placed into a tissue culture dish (100x20 mm), cleaned by removing the faecal material and connective tissues, and rinsed by cold PBS. The ileum sample was then cut into small pieces, transferred into a 50 mL centrifuge tube with 10 mL 1x IEL buffer (listed in appendix A2) and incubated at 37°C for 30 minutes with rotation at 220 rpm for complete digestion. The whole digestion product was slowly poured into a BD Falcon cell strainer (40 μm) placed on a 50 mL centrifuge tube to allow the single cells to drop into the tube. The cells were then washed twice with PBS (containing 2% FBS) by centrifuging at 4°C for 10 minutes (200x g) and counted using a hemocytometer under the microscope. If many clumps were observed, the cells were separated by using syringe with 21G needle to expel the cell solution. Finally the appropriate amount of cells were divided into two tubes

and resuspended with EDTA-PBS buffer (PBS, 2%FBS, 5 mM EDTA) and DNase I-PBS buffer (PBS with calcium and magnesium, 2% FBS, 0.1 mg/ml DNaseI) respectively for flow cytometry sorting.

3.2.3 PCR protocol using H2 primers

The cDNA templates were generated always following the protocol in section 2.2.2 chapter 2. A standard PCR protocol (Table 3.1 & 3.2) using primers called H2, H2-PCR, was established by Dr. John Chattaway to amplify haemopoietic BG genes with type 2 3' UTR sequences.

3.2.4 Primer design

The primers used for sequencing purposes were designed using online programme 'Primer3Plus' (http://primer3plus.com/). The new primers aiming to amplify all haemopoietic BG genes were designed manually following the principles described in section 2.2.3 chapter 2.

Table 3.1 The H2-PCR recipes

PCR reagents	
Nuclease free water	32.5 µL
10 mM total dNTP (2.5 mM each)	1 µL
Primer UC206 (10 µM)	2 µL
Primer UC210 (10 µM)	2 µL
5x Phusion buffer (NEB)	10 µL
Phusion enzyme (NEB)	0.5 µL
Total:	50 µL

Table 3.2 The H2-PCR programmes

95°C	2 min	
95°C	45 s	5 cycles
68.4°C	30 s	
72°C	90 s	
95°C	45 s	30 cycles
68.4°C	1 s	
72°C	90 s	
72°C	10 min	
10°C	~	

3.2.5 Ligation and colony PCR

TOPO ligation and pJET ligation kits were both tested in this chapter to ligate PCR products into respective vectors from the kits for the purpose of getting colonies carrying BG gene sequenced. Both ligation kits worked well and pJET was chosen to apply to the whole project for its cost-effectiveness.

Colony PCR was done primarily following the protocol described in section 2.2.10 chapter 2, but the method for preparing colony PCR templates was different when performed on B cells sample from line N (B21), for which 94 colonies were picked up using HU primer pair. To handle such large number of colonies, sterile 96 well plates with 30 µl LB buffer in each well were used to maintain the original colonies being picked. Then the plates were placed into 37°C incubator for 2 hours, and afterwards, 2.5 µl cultures from each well were transferred into 96 well PCR plate using multi-pipet for PCR reaction.

3.2.6 Sequence analyzing tools

Sequencing trace data was viewed in CLC DNA Workbench 5 (CLC for short), and only reliable trace data were applied to assemble. The assembly programmes from CLC and

SeqMan (DNASTAR®) were compared and SeqMan was chosen as it was more flexible for manual operations, for example the parametre setting changes etc., to fit the experimental purpose. And once the assembling was completed, the consensus sequence generated by SeqMan was exported as a single sequence and imported into CLC for further analysis, e.g. sequence alignment, protein translation etc.

3.2.7 Sequence annotation

The gene structures of cDNA sequences were determined by aligning and comparing them to the well annotated BG genomic sequences of B12 haplotype chicken (GenBank: KC955130.1) using CLC. BG genes share similar intron/exon structure giving a full RNA with a relatively long 5' UTR (164-312 nt) and 3' UTR (around 470 nt), and in between a coding sequence to a signal peptide, a Ig-like V domain, a transmembrane region, and a cytoplasmic tail composed of many heptad repeats. Among the structures above, the same sizes and very close nucleotide compositions of Ig-like V domain and transmembrane region were shared by all BG genes found so far; the cytoplasmic tails were formed primarily by various numbers of heptad repeats, leading to the idea that two cytoplasmic tails form a coiled coil. Therefore, by comparing the cDNA sequences to BG genomic genes of B12 haplotype, the structures above could be separated and fitted into equivalent regions on the genomic template automatically, and then could be annotated with manual check and changes.

3.2.8 Phylogeny study

DNA or protein sequences were first aligned using CLC and exported into a local folder as an '.aln' file. Then the alignment file was opened in MEGA7 (Molecular Evolutionary Genetics Analysis, http://www.megasoftware.net/) software, realigned again using ClustalW (Thompson *et al.*, 1994; Larkin *et al.*, 2007) for DNA sequences or MUSCLE (Edgar, 2004a; 2004b) for protein sequences without changing the default parameter settings, and the alignment result was exported into a local folder as an '.meg' file. Finally the '.meg' file was

applied to construct a phylogenetic tree using Neighbor Joining (Saitou *et al.*, 1987) method in MEGA7 with the bootstrap replicates set to 1000.

3.2.8.1 ClustalW and MUSCLE

ClustalW and MUSCLE were the only two alignment methods offered in MEGA7 but were sufficient for the project. ClustalW was widely used for decades especially in the multiple DNA sequences alignment (Hung *et al.*, 2016). In this project, no default setting was changed but when the DNA sequences were encoding Ig-like V domain, transmembrane or coiled coil, the answer to alignment choice was 'Align Codons' instead of 'Align DNA'. Because the three types of sequences above were protein coding sequences, and these sequences were ensured to be aligned by codons under the choice of 'Align Codons', which would provide a much more realistic approach than direct alignment through avoiding introducing gaps into positions. MUSCLE, stands for Multiple Sequence Comparison by Log-Expectation, was used for protein sequence alignment because it was claimed to achieve better average accuracy in multiple protein sequences alignment compared to ClustalW (claim in www.ebi.ac.uk/Tools/msa/muscle/help/index.html).

3.2.8.2 Neighbour Joining

Neighbour Joining (NJ) is one of the distance-based methods that analyses clusters and determines minimal distances between sequences. NJ is now the most used method for constructing phylogenetic trees and computing lengths of each branches in the tree (Mailund *et al.*, 2006). It was suggested that combination of NJ and bootstrap (Felsenstein, 1985) analysis might be the best way to evaluate trees using distance methods (Nei *et al*., 1998). In MEGA7, the bootstrap can be added to evaluate NJ trees by changing the default parametre settings in the option window of 'analysis preferences'. Under the option window, bootstrap method was chosen for the phylogeny test, and 1000 bootstrap repeats was answered for the question of 'No. of bootstrap replications' for this project.

This page intentionally left blank

3.3 Results

3.3.1 Quality of purified RNA from sorted cells was sufficient to satisfy the purpose of this project

In order to understand BG expression at RNA level from specific cell types (T cells, B cells and IECs), cells sorted either by our collaborators or in the department were tested for RNA quality by PCR amplification of the housekeeping gene GAPDH. All PCR results confirmed the RNA integrity. The PCR for sorted T cells and B cells from B19 and B21 haplotype chickens were carried out by our collaborator Ms. Karen Staines (data not shown). The PCRs for IECs sorted from B19 and B21 haplotype chickens were done by myself and the results were shown in figure 3.2E.

For the IECs sorted from villi of ileum, two different reagents for sample treatment were compared, and both FACS and PCR results showed little difference between the two treatment (Figure 3.2). According to previous sorting experience from Dr. Michaela Fakiola, cells were often found as clumps which were difficult for sorting single cells. Both EDTA and DNAse I were reported to be effective for preventing cell-cell clumping during sample preparation. However, these two reagents function in totally different conditions: DNase I requires magnesium solution while EDTA has a high affinity for metal cations. Therefore, two different buffers ('EDTA treat' buffer and 'DNase treat' buffer), were tested independently to treat cell preparation before antibody staining, and they were found to be equally effective by the following PCR results (Figure 3.2E and F).

Both the RNA quality (Figure 3.2 E) and the sorting quality (Figure 3.2 F) were evaluated by two different PCRs. In one PCR, the quality of RNA integrity of sorted IECs were confirmed by large yields of GAPDH gene products. In the other PCR, primers specifically amplifying CD45 marker, a marker for haemopoietic cells, were used to indicate the purity of IECs after sorting. If IECs were pure from contamination of other types of cells, there should be no band found in the PCR reaction. Through comparing the PCR results between unsorted cells and

sorted IECs, the sorted IECs from both two different treatment above were confirmed to be pure enough to satisfy the purpose of this project.

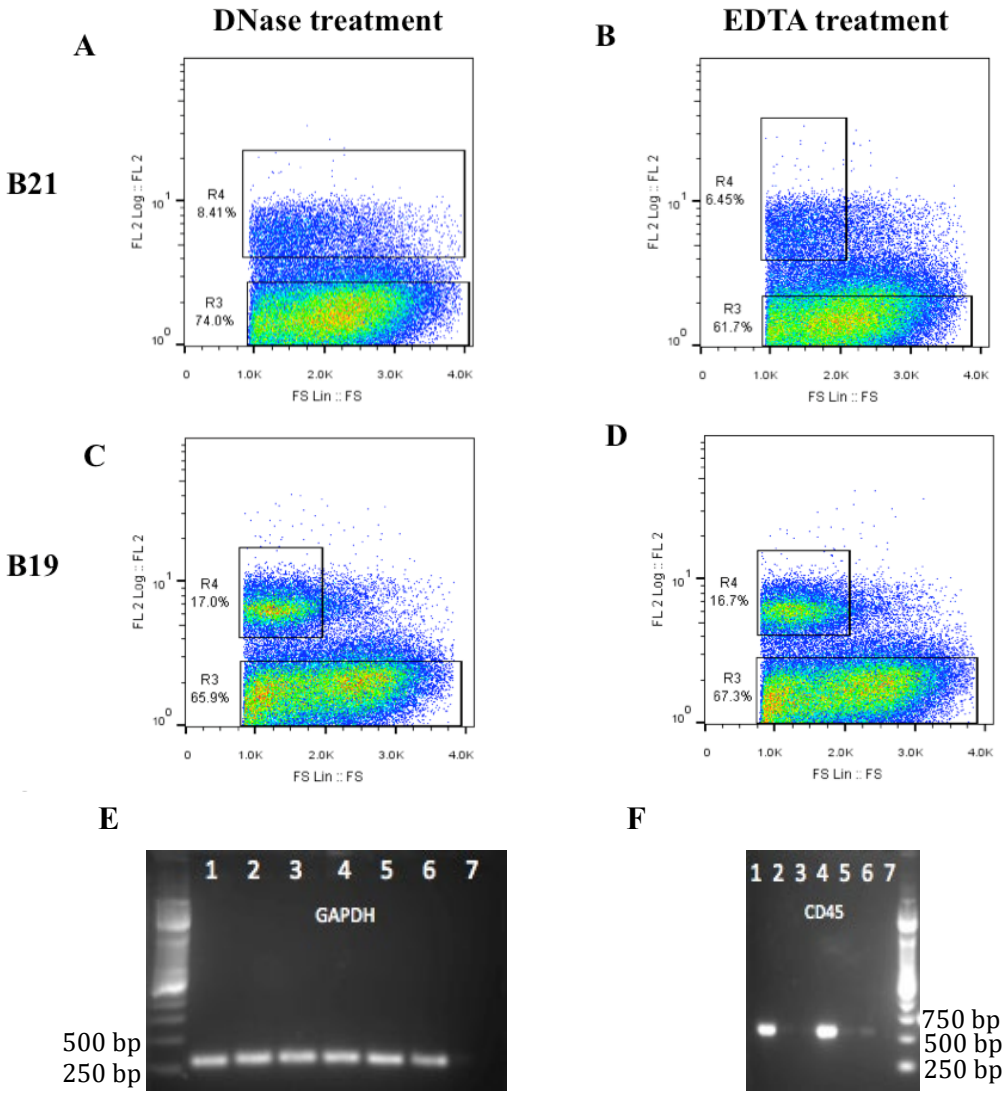


Figure 3.2 The intestinal epithelial cells (IECs) sorted from B19 and B21 haplotype chickens. FACS results are shown in A-D with gating on live cells. The x-axis is forward scatter showing the cell sizes and the y-axis was cells stained with PE conjugated anti-CD45 antibody. IECs were collected from gates R3. A and B are samples from B21 haplotype treated with DNase I and EDTA, respectively; C and D are samples from B19 haplotype treated with DNase I and EDTA, respectively. Post-sorted IECs were amplified by PCRs to confirm the RNA integrity (Fig E) and the sorting quality (Fig F). Seven lanes in figure E and F refer to cDNA from: Lane 1. N line pre-sorting cells; Lane 2, N line post-sorting cells treated with DNase I; Lane 3, N line post-sorting cells treated with EDTA; Lane 4, P line pre-sorting cells; Lane 5, P line post-sorting cells treated with DNase I; Lane 6, N line post-sorting cells treated with EDTA; Lane 7, negative control using nucleotide-free water as template.

3.3.2 H2 primers worked on tissue sample but not on sorted cells

3.3.2.1 A specific band from duodenum sample was seen on DNA electrophoresis gel

The duodenum sample, a chunk of the whole tissue, from B15 haplotype chicken was tested for BG expression on RNA level by two independent PCRs following Chattaway's H2-PCR protocol described in section 3.2.3. There were obvious bands found in the DNA gel electrophoresis with about the right size around 2000 bp (Figure 3.3). The bands were cut and cloned separately for sequencing to confirm whether they belong to BG genes.

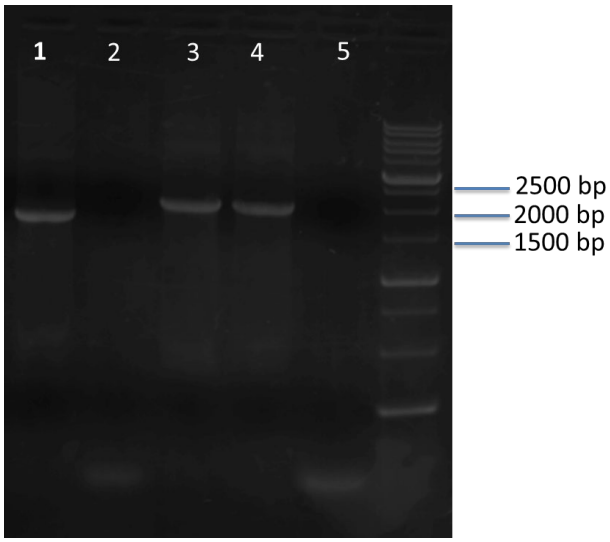


Figure 3.3 The DNA electrophoresis gel of PCR amplification of BG genes from duodenum sample of B15 haplotype chicken using H2 primers. Lane 1. positive control using plasmid 'clone K751 (B12-PCR1-27)' (provided by Dr. John Chattway) as template in the 1st independent H2-PCR reaction; Lane 2, negative control using nucleotide-free water as template in the 1st independent H2-PCR; Lane 3, the 1st H2-PCR reaction using cDNA template from duodenum sample of B15 haplotype; Lane 4, the 2nd H2-PCR reaction using cDNA template from duodenum sample of B15 haplotype; Lane 5, negative control using nucleotide-free water as template in the 2nd independent H2-PCR.

3.3.2.2 PCR bands were faint on sorted T and B cells

The sorted T and B cells from both spleen and blood samples of both B19 and B21 haplotype chickens were also tested for BG gene expression using the same H2-PCR protocol as that used above for duodenum sample in section 3.3.2.1; however there was no obvious band found in these PCRs. The quality of cDNA was confirmed by PCR amplification of housekeeping gene GAPDH (data not shown). The cDNA was applied in seven attempts to amplify BG genes using H2 primers but each time there were only several very faint bands showed in DNA gel electrophoresis with different sizes. Then the PCR reaction was adjusted by changing PCR conditions, e.g. the annealing temperature, as well as using new Fusion kits. However, each time there were still very faint bands observed. In order to understand what these faint bands were, the PCR products from several PCR reactions were concentrated using ethanol precipitation, followed by DNA gel electrophoresis, and a large area of gel containing these faint bands was cut and purified. Only trace DNA products were isolated which failed further cloning.

Two possible reasons were considered for this failure. First, H2 primers were designed based on BG genes mainly from B12 haplotype, therefore, they might not be able to bind to BG gene template from neither T cells nor B cells of B19 and B21 haplotype due to the polymorphism. Though several 5' UTR and 3' UTR sequences from B21 haplotype found by Chattway were aligned with B12 BG sequences when designing H2 primers (Figure 3.1), the BG genes in T cells and B cells from B19 and B21 haplotypes might differ to these sequences used in the alignment. In this case, new primers needed to be designed. Second, there was no BG gene expressed in T or B cells of the two haplotypes above, or the gene expression level was too low to be detected by PCR. However, this possibility had to be tested by either designing new primers to repeat PCR or using other methods, for example western blot to confirm the BG protein expression.

3.3.2.3 There was no PCR band found on sorted intestine epithelial cells (IECs)

The sorted IECs from B19 and B21 haplotype chickens were also examed for BG expression using the same H2-PCR protocol as that used above for duodenum sample in section 3.3.2.1, and no band was observed in DNA gel electrophoresis at all. As shown in figure 3.2 in section 3.3.1, PCRs were used to test cDNA integrity by amplifying GAPDH gene, therefore, it could be concluded that either there was no BG expression in IECs from B19 and B21 haplotypes, or the IEC BG genes could not be amplified by H2 primers. The latter explanation is more likely as IECs are tissue cells, while H2 primers are targeting haemopoietic BG genes.

3.3.3 Internal sequencing primers were designed to get full length of BG cDNA sequences

The colonies confirmed to contain BG genes by colony PCR from section 3.3.3 were sequenced using T7 forward and pJET reverse primers. However, the results of two sequencing runs could not be assembled as the two primers above did not yield reliable sequences in the middle of the DNA clones. The average length of BG cDNA was around 2000 bp according to B12 haplotype (Salomonsen *et al.*, 2014), while 400 – 800 bp was the normal read length with reliable trace data provided by Sanger sequencing facility here in Department of Biochemistry, University of Cambridge. Therefore, internal sequencing primers were needed to complete the whole sequence.

Internal sequencing primers designed both by John previously and by myself were tested, and all the primers performed well in the PCR amplification, but when they were used in sequencing, ‘mixed peaks’ (Figure 3.4) frequently appeared in the sequencing data. ‘Mixed peaks’ might be caused by primers binding to more than one site, especially for primers targeting on cytoplasmic tail region where the nucleotide sequence template had repeats in other sites of the cytoplasmic tail region. Also, due to the polymorphism of BG genes, some internal sequencing primers worked well on some haplotype might not work well on other

haplotype. In this chapter different internal sequencing primers were applied to BG sequences from different haplotypes with the information listed in table 3.3.

Table 3.3 The internal forward and reverse primers for sequencing BG colonies

Primer ID	Primer	haplotype	Designer
UC453	5' TGTKGTGYTGYGCTGCCA 3'	B15	Chattaway
UC647	5' GCAGTGTTCTCACTCAA 3'	B21	Chen
UC649	5' GCAGTTMATHTCTCARTC 3'	B21	Chen
UC701	5' TGGCTCTGCACY(C/T)TCCTCS(C/G) 3'	B21	Chen

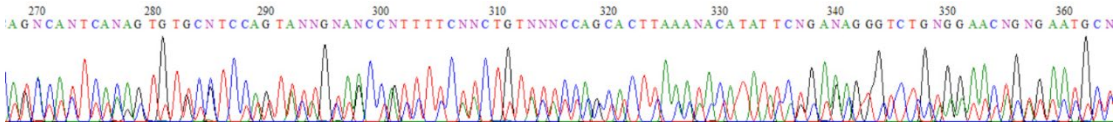


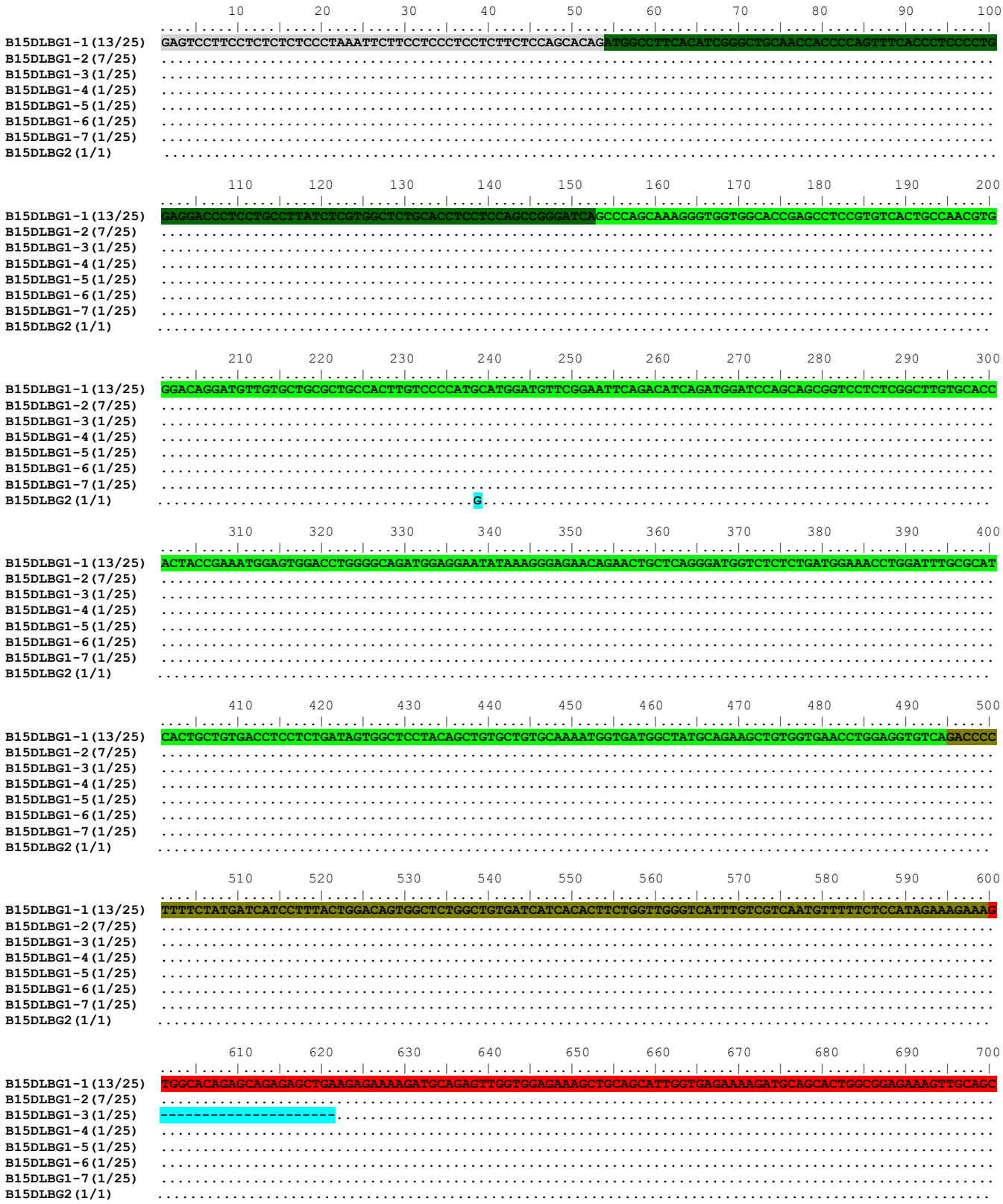
Figure 3.4 An example of "mixed peaks" sequencing report from one DL-B15 clone.

3.3.4 Two new BG genes were found in B15 duodenum using H2 pairs

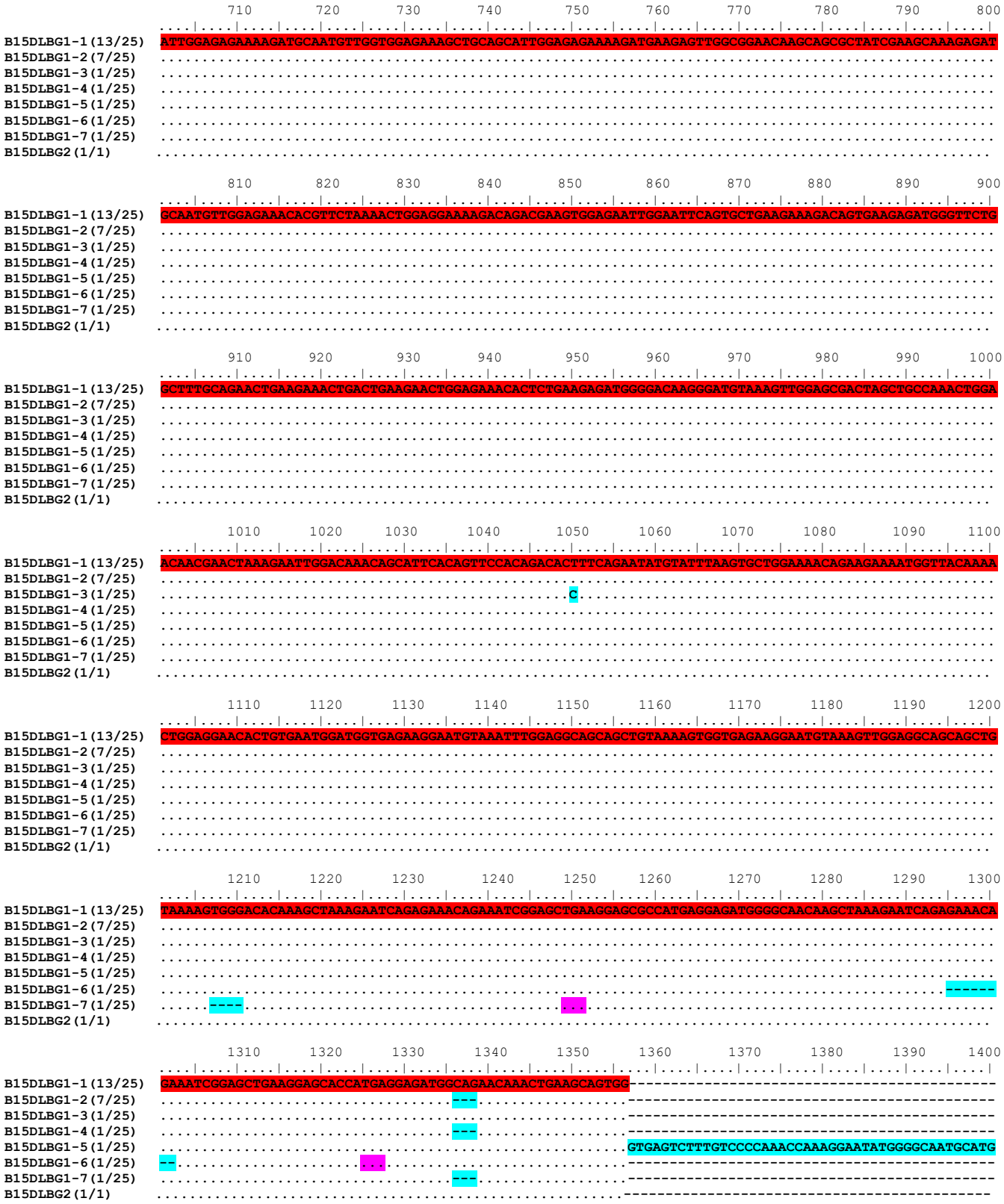
Two independent PCR reactions with the same cDNA template using H2 primers on duodenum tissue of B15 haplotype were performed to investigate BG gene expression at the cDNA level. In each experiment, PCR products were cloned into pJET vector, followed by transformation. In total 12 colonies from the first independent PCR and 14 colonies from the second independent PCR, which had been confirmed to have a DNA insertion by colony PCR, were grown up, made miniprep and sequenced. Sequencing results showed that two new BG genes were found: one new BG gene was found in both independent PCR reactions with alternative splicing in the cytoplasmic tail region while the other new potential BG gene was only found once in the 2nd PCR reaction with only one colony (Figure 3.5).

3.3.4.1 Nomenclature of new BG gene found from B15 duodenum sample

In this project, the two new BG genes were named B15DLBG1 and B15DLBG2 respectively, with ‘B15’ for B15 haplotype, ‘DL’ for duodenum loop, ‘BG1’ for the BG gene found with most clones, and ‘BG2’ for the other BG gene with only one clone. There were seven alternatively spliced isoforms found under B15DLBG1, and they were named with their gene named followed by a dash, the isoform number, and number of colonies that showed the same sequence. For example, ‘B15DLBG1-2(x7)’ and ‘B15DLBG1-2(7/25)’ both reflect the same sequence: the 2nd isoform of BG1 cDNA sequences found in duodenum sample of B15 haplotype with 7 colonies showing the same sequence. The difference in latter name was that the whole colony number showing BG1 cDNA sequences (25 colonies) was indicated as well.



The figure legend is shown on the page following the figure



The figure legend is shown on the page following the figure

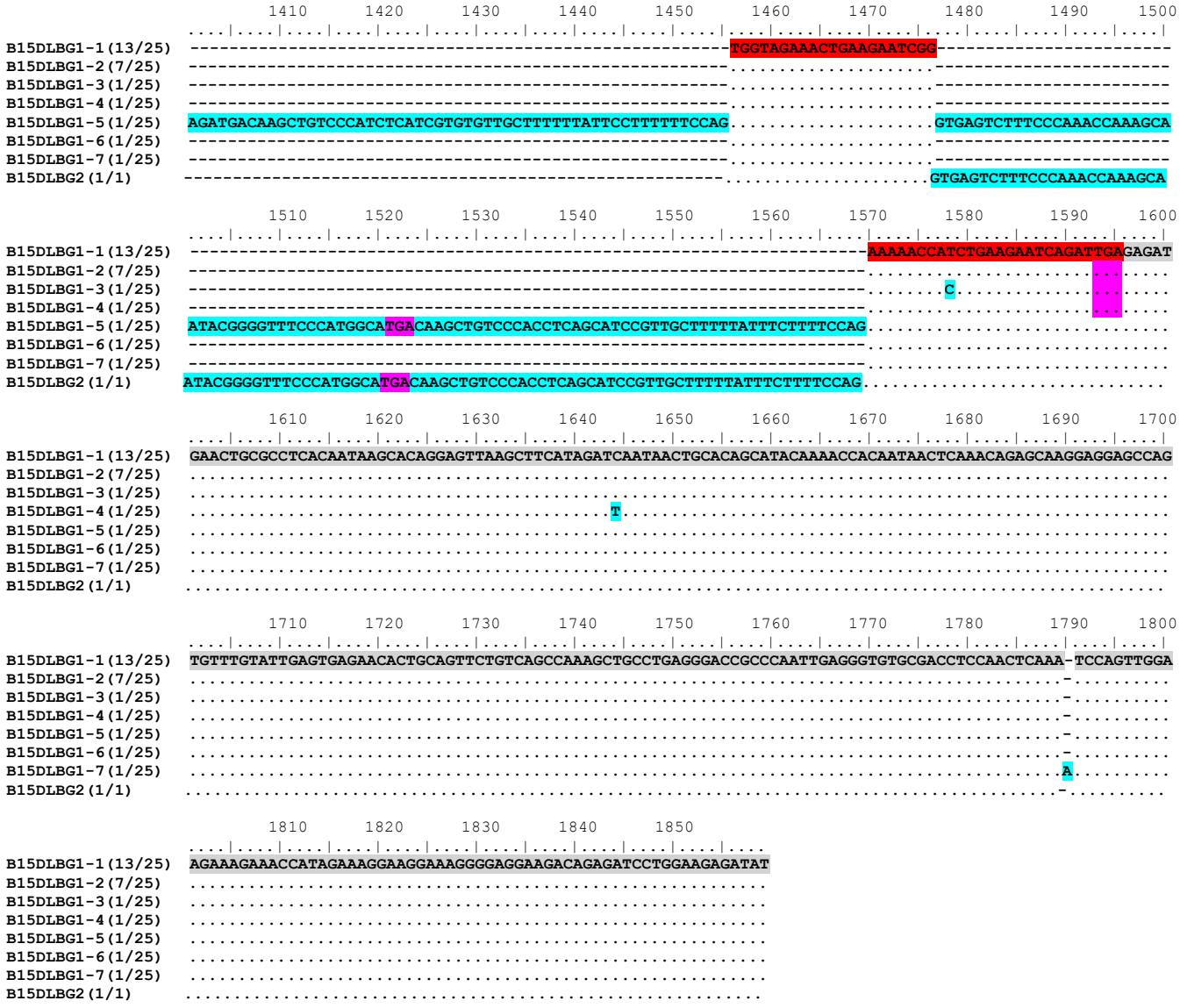


Figure 3.5 The nucleotide sequence alignment of all BG sequences found in duodenum sample of B15 haplotype using H2 primers. The top seven sequences are alternative splicing isoforms of B15DLBG1 gene with their colony numbers found plus the total number reflecting whole B15DLBG1 gene. And the bottom sequence is B15DLBG2 gene which was only found with one clone. Regions of the nucleotide sequence are indicated with colours: 5' UTR and 3' UTR, grey; signal sequence, darker green; Ig-V domain, bright green; transmembrane region, darker brown; cytoplasmic tail region, red. The sites labeled by blue in other sequences indicate differences compared to the top sequence. The stopping codons are labeled in purple for all the sequences except the top one. Letters indicate nucleotides, dots indicate identities with the top sequence, dashes indicate no sequence present compared to one or more of the other sequences.

3.3.4.2 There were seven alternative splicing cDNA transcripts in B15DLBG1

As shown in figure 3.5, there were seven alternatively spliced cDNA transcripts found from the B15DLBG1. To better understand the relationships between these transcripts and ensure the alignments are accurate for further phylogenetic studies, all the sequences were trimmed to start from the same position in 5' UTRs and end in the same position in 3' UTRs. The sequence lengths from B15DLBG1-1 to B15DLBG1-7 shown in figure 3.5 were 1666 bp, 1663 bp, 1645 bp, 1663 bp, 1858 bp, 1658 bp and 1660 bp, respectively. The major difference in sequence lengths was due to the alternative splicing in the cytoplasmic tail region, while no difference was found in their 5' UTR regions, Ig-V domains or transmembrane regions. In theory, these cDNA sequences might encode BG proteins with the same Ig-V domain and transmembrane region but different cytoplasmic tails (Figure 3.6).

Compared to B15DLBG1-1, the dominatingly expressed cDNA sequence, all the differences on other isoforms were summarized in table 3.4. B15DLBG1-2 is almost virtually identical to B15DLBG1-1 except for one trinucleotide, which corresponds to an alternative splice site, three nucleotides before the other splices, a situation already been found in other BG heptad exons (Chattaway, 2013) and also for the isoforms of B15DLBG1-4 and B15DLBG1-7. The third isoform, B15DLBG1-3, has a deletion of 21 nucleotides (a putative exon) resulted from alternative splicing. B15DLBG1-5 has two insertions both start with GT and end with AG, indicating unspliced introns. The first insertion (99 nt) could not be translated into a full protein sequence due to a stop codon in the middle, while the second insertion (93 nt) could be translated correctly. Such phenomena were observed previously in B19 cDNA and were considered as unspliced variants (Kaufman et al., 1990). B15DLBG1-6 and B15DLBG1-7 both have a deletion, but with 8 nucleotides and 4 nucleotides, respectively, which we do not understand so far. There is also one nucleotide mutation occurring in three isoforms, B15DLBG1-3, -4 and -7, which are likely to be PCR errors. However, except the two major isoforms, B15DLBG1-1 and -2, all other isoforms are presented only once, therefore, we could only consider the two major isoforms valid. More discussion regarding to alternative splicing and its mechanisms were discussed in chapter 4 and 5.

Table 3.4 The differences among all 7 B15DLBG1 isoforms

Sequence ID	colony No.	Differences	Sites (bp)
B15DLBG1-1 (13/25)	1st : 7; 2nd : 6	none	
B15DLBG1-2(7/25)	1st : 3; 2nd : 4	3 nts del	1336~1338
B15DLBG1-3(1/25)	1st : 1	21 nts del, 1 nt mut	1578
B15DLBG1-4(1/25)	1st : 1	3 nts del, 1nt mut	1336~1338, 1644
B15DLBG1-5(1/25)	2nd : 1	99 nts 93 nts ins	1359~1457, 1477~1569
B15DLBG1-6(1/25)	2nd : 1	8 nts del	1295~1302
B15DLBG1-7(1/25)	2nd : 1	4 and 3 nts del, 1nt ins	1207~1210, 1336~1338, 1790

Note: '1st'/'2nd' for 1st /2nd time PCR; 'del' for deletion; 'mut' for mutation; 'ins' for insertion.

3.3.4.3 There was only one nucleotide difference in V domain between the two new genes

Apart from the difference in the cytoplasmic tail region, which looks like intron read-through, B15DLBG1 and B15DLBG2 differ in only one nucleotide found in the Ig-like V domain resulting in a different amino acid (Figure 3.5). The difference in V domain was considered as the major rule distinguishing different BG genes in this project. In this experiment, B15DLBG2 gene was only found once in one PCR reaction with only one colony, so it seems likely that B15DLBG2 is the result of a simple misincorporation during the PCR. However, a similar phenomenon was observed between two genes, BG8 and BG9 in B12 haplotype chicken, of which only a few nucleotide differences occurred in 5' UTR and V domain, and a little difference in cytoplasmic tail region. Therefore, to really understand whether B15DLBG2 was a true new BG gene or was just a PCR artifact, another independent PCR needs to be done to pick up more colonies for sequencing, or genomic sequencing should be carried out.

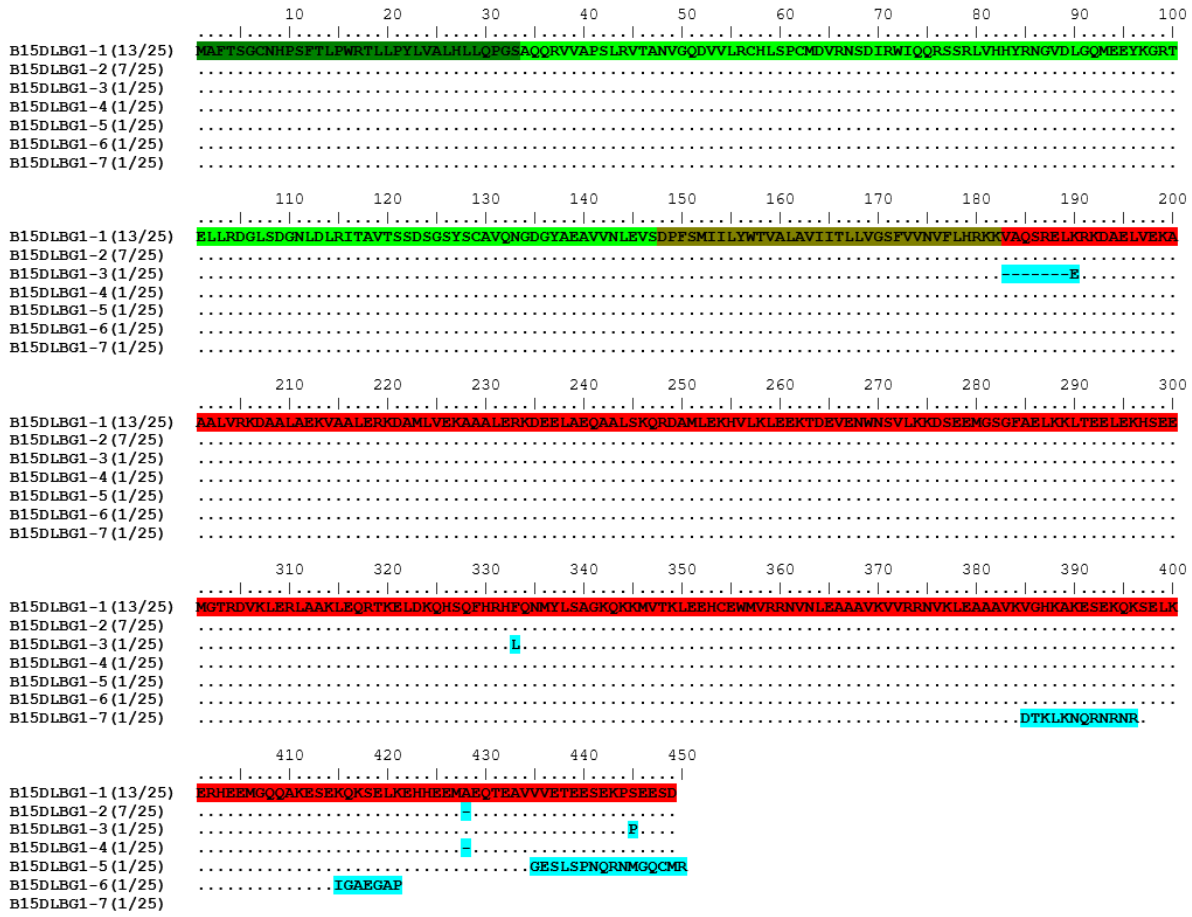


Figure 3.6 The amino acid sequence alignment of all B15DLBG1 alternative splicing transcripts. The nomenclature of each alternative splicing isoform on the left side is followed by gene name, isoform identification (ID) number and the colony number showing this isoform sequence plus with totally colony numbers of this gene. Take the top sequence's name 'B15DLBG1-1(13/25)' for example, it can be understood that this is first alternative splicing isoform of BG gene B15DLBG1, and 13 colonies among 25 in total showed this sequence. Regions of the amino acid sequence are indicated with colours: signal sequence, darker green; Ig-V domain, bright green; transmembrane region, darker brown; cytoplasmic tail region, red. The sites labeled by blue in other sequences indicate differences compared to the top sequence. Letters indicate amino acids by single letter code, dots indicate identities with the top sequence, dashes indicate no sequence present (deletion) compared to the top sequence.

3.3.5 Phylogenetic study showed the two new B15 BG genes belong to BG8-9-12 group in B12 haplotype

To understand the phylogenetic relationship between the two new BG genes found in duodenum sample of B15 haplotype chicken and 14 BG genes annotated from B12 haplotype chicken, nucleotide sequence alignment as well as phylogenetic trees were built. Firstly, only the dominantly expressed alternatively spliced isoform from B15DLBG1 gene (B15DLBG1-1) was chosen to make the alignment with B15DLBG2 and 14 BG genes of B12 haplotype (BG-B12). The two new genes contained shorter 5' UTR and 3' UTR sequences compared to that of 14 BG-B12 genes due to the trimming of bad sequencing quality on the primer regions; therefore, the extra sequences of these two regions on B12 BG genes were deleted in order to make the alignment and phylogenetic analysis more reliable and accurate (Appendix C). Then phylogenetic trees were built both based on the whole cDNA sequence alignment as well as the individual regions aligned separately (Figure 3.7).

By comparing the phylogenetic trees here to the trees made only with the 14 BG genes of B12 haplotype (Salomonsen *et al.*, 2014), it could be seen that the trees were not disturbed by introducing the two new BG genes from B15 haplotype, and the two new genes were always grouped into BG8, 9 and 12 clusters. In the tree made by whole cDNA sequences, both B15DLBG1 and B15DLBG2 were obviously grouped into the BG8, 9 and 12 cluster, indicating they were evolutionarily close to these three genes. The 5' UTRs of the two new genes fell into the haemopoietic BG gene clusters, indicating they were haemopoietic BG genes. Although they were from a duodenum chunk, blood was likely contained in this sample, or they might be from lymphocytes existing in the lamina propria. The 3' UTR tree proved that the 3' UTRs of these two new genes belonged to type 2. The two new genes were amplified by H2 primers which was exactly designed to amplify haemopoietic BG genes with type 2 3' UTRs; in B12 haplotype, the cluster that has 'H2' was BG8, 9, 12 and 13 (Salomonsen *et al.*, 2014). From all the trees in figure 3.7, B15DLBG1 and B15DLBG2 sometimes were closer to BG9 and sometimes closer to BG8, but in general, they were within

the BG8, 9 and 12 group. The BG8, 9 and 12 group, as described in section 3.3.3.4, has only a few nucleotide differences between these genes. Therefore, BG8, 9, 12 and the two new BG sequences found from B15 haplotype (BGB15DLBG1 and B15DLBG2) might be evolved from the same ancestor, and the different copies were due to the duplication event.

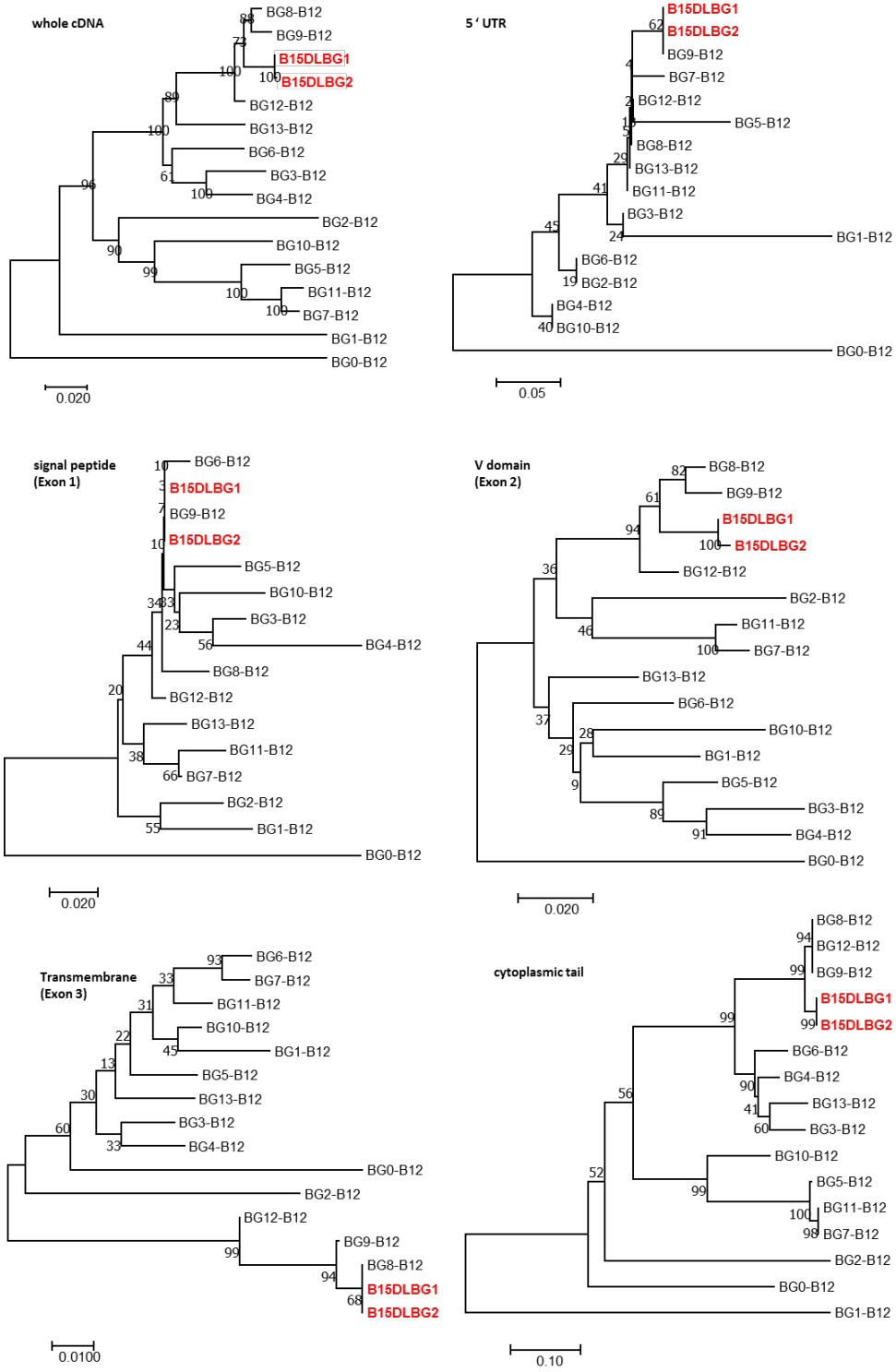


Figure legend is shown on the page following the figure

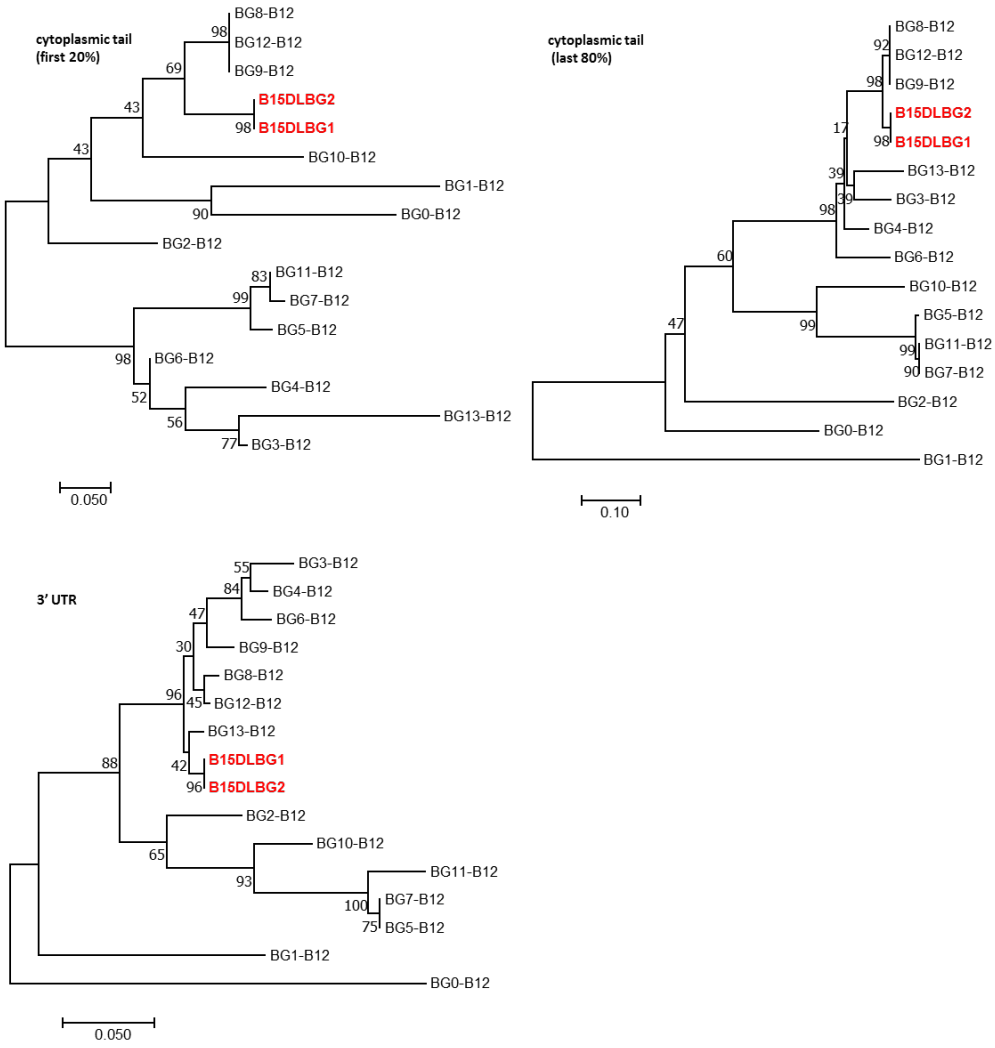


Figure 3.7 Phylogenetic analysis of nucleotide sequences between two new BG genes from duodenum sample of B15 haplotype and 14 BG genes of B12 haplotype. Phylogenetic trees built on full length sequences and individual regions all show that the two new BG genes labelled in red are clustered into BG8, 9 and 12 group, except the trees built on signal sequences and 3' UTRs. In the 3' UTR tree, the new genes are closer to BG13, while in the signal sequences tree, no obvious relationship is observed.

3.3.6 A new primer pair HU was designed to amplify all haemopoietic BG genes

As discussed previously, the H2 primer pair, designed to amplify haemopoietic BG genes with type 2 3' UTR sequences, was inefficient for PCR in sorted T cells, B cells and IECs samples from B19 and B21 haplotype chickens (section 3.3.2). Therefore, new primers need to be designed. Ideally, a pair of universal primers which could amplify all haemopoietic BG genes with both type 1 and type 2 3' UTRs could firstly simplify the process of examining BG genes from T and B cells, and secondly satisfy the purpose of evaluating which BG was dominantly expressed in T and B cells. To design the new universal primers, all the 5' UTR sequences and 3' UTR sequences of BG genes available so far, including 14 BG genes from B12 haplotype, annotated BG genes from genomic sequence of BQ haplotype, and the cDNA sequences from our previous work done by Dr. Chattaway, were aligned to design primers.

Because the primer pair had to be set at the end of 5' UTR and 3' UTR in order to get as much the whole cDNA sequence as possible, and due to the high polymorphism of BG genes, no primer was recommended by any primer designing tool. The reasons given by these tools were either the Tm was too high, or secondary structures (dimers and hairpins) would occur. As a consequence, a degenerated reverse primer called 'U' for universal (with the ID of UC650), based on all the BG sequences with both type 1 and type 2 3' UTRs accessible was designed manually, which was conserved with both haemopoietic and tissue BG genes (Figure 3.8). As for the forward primer, since primer H (ID number: UC206) was in the very end of 5' UTR and conserved in haemopoietic BG genes, it was used as the forward primer. Therefore, the new primer pair (UC206 and UC650) was named HU primers or HU primer pair, which was then optimized to amplify all haemopoietic BG genes with both type 1 and type 2 3' UTRs.

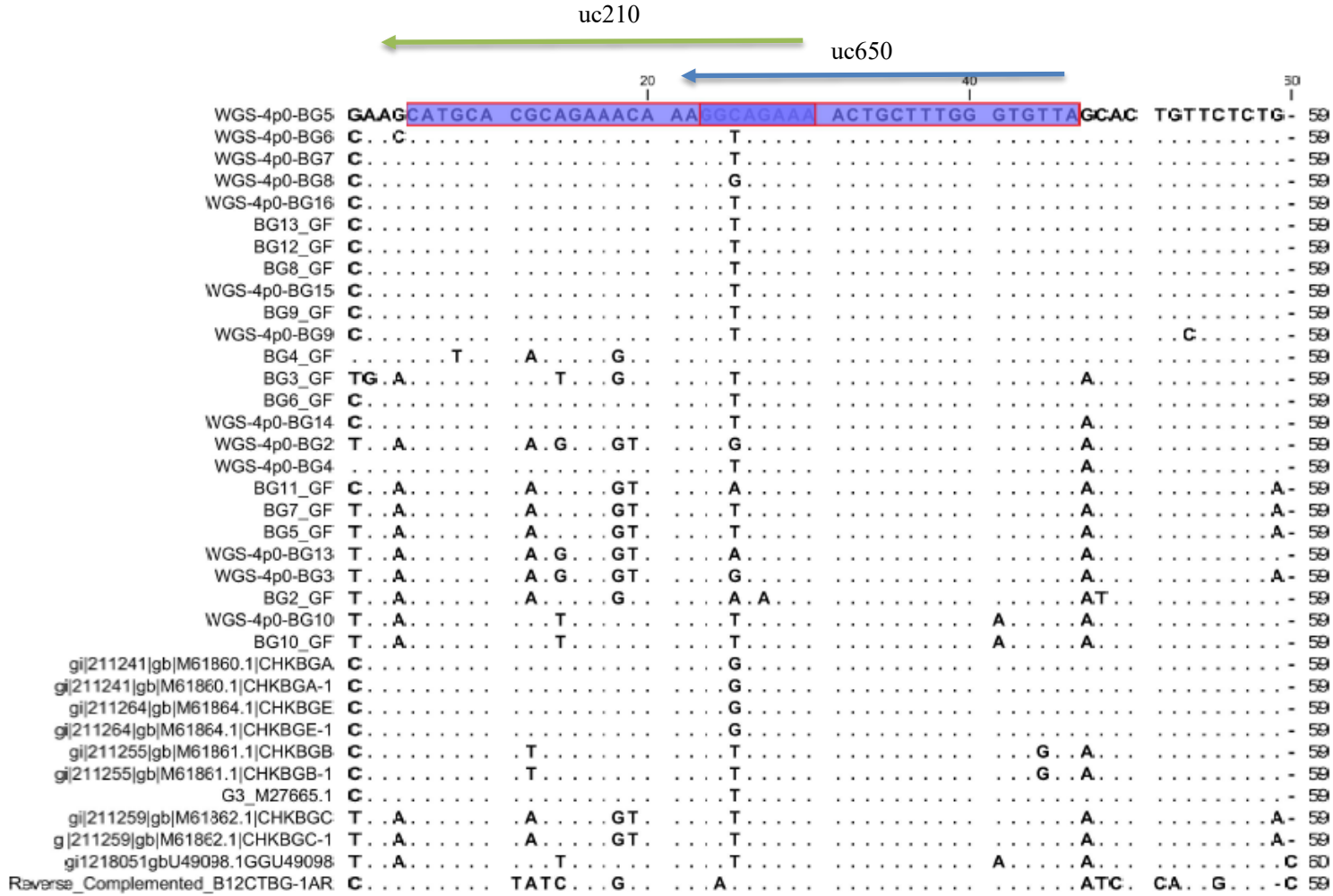


Figure 3.8 Comparison of sequence locations of reverse primers in H2 and HU primer pairs. The nucleotide sequence alignment was created based on the end of 3' UTR sequences of all BG cDNA sequences available in 2015, and the locations of reverse primer uc206 in H2 primer pair and uc650 in HU primer pair are indicated by green arrow and blue arrow respectively. The oligo information is listed in appendix B.

3.3.7 HU-PCR optimization

The efficiency of amplifying BG genes by the HU primer pair was first tested using cG3 cosmid template made from B12 haplotype (Salomonsen *et al.*, 2014) following similar recipe and the same PCR programme as H2-PCR reaction. However, when using cDNA generated from sorted B cells of line N (B21 haplotype) chicken as template, apart from the expected band (around 2000 bp), there were many unexpected bands on DNA gel electrophoresis despite many attempts of optimization (Figure 3.10). For example, the PCR annealing temperature were tested by gradient PCR, annealing and extension time were adjusted, and even several different PCR kits were tried.

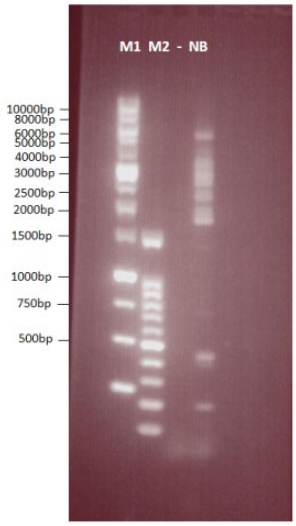


Figure 3.9 The DNA electrophoresis gel of PCR amplification of BG genes from sorted B cells of line N (B21) chicken using HU primer pair. Lane M1, 10 kb DNA marker; Lane M2, 1 kb DNA marker; Lane 3, negative control using nucleotide-free water as template; Lane 4, cDNA template from sorted B cells of B21 haplotype.

In order to understand what these bands were, PCR bands around 1500 bp, 2000 bp, 2500 bp, 3000 bp and 4000 bp were cut and purified separately to get the DNA fragments which then were cloned into pJET vector. In total, nighty four colonies were picked for colony PCR, all 46 colony PCR positive clones were done with miniprep and sent for sequencing, which the result showed that 4 colonies were not BG sequences, 12 colonies were genomic BG sequences with size of more than 3000 bp, and 30 colonies were BG cDNA sequences with alternative splicing.

3.3.8 A method was developed to simplify the cloning procedure

As all the DNA bands ranging from 1500 bp to 3000 bp produced by HU-PCR were BG sequences with alternative splicing, the HU-PCR optimization could be considered successful. However, the way of cutting bands and doing cloning separately was time and labour consuming which would make the project impractical. To solve this problem, a method was developed to simplify the cloning procedure.

After the HU-PCR reaction, all the PCR products were loaded to run the DNA gel electrophoresis with much shorter time (20 min at 100 V) compared to normal DNA gel electrophoresis (60 min at 100 V). Under such circumstance, the DNA bands could not be completely separated and the bands with size ranging 1500 bp to 3000 bp could be easily cut in one small piece of gel, which was easier for DNA purification. DNA fragment mixtures were cloned into pJET vector, and 96 colonies were picked for colony PCR following the protocol described in section 3.2.5. All the colony PCR positive clones were done by miniprep and sent for sequencing. The sequencing results showed the same as the method of cutting DNA bands separately. Thus an efficient method to investigate all the BG sequences was developed to examine BG expression using HU-PCR reaction.

3.3.9 Standard protocol for investigating haemopoietic BG genes using HU primers

To summarize, a standard protocol for investigating haemopoietic BG genes using HU primers was developed and explained as follows.

3.3.9.1 HU-PCR protocol

After optimization, the standard HU-PCR recipe and programme were listed in table 3.5 and 3.6.

Table 3.5 The HU-PCR recipes

PCR reagents	
Nuclease free water	35.5 µL
10 mM total dNTP (2.5 mM each)	1 µL
Primer UC206 (10 µM)	1.25 µL
Primer UC650 (10 µM)	1.25 µL
5x Phusion buffer (NEB)	10 µL
Phusion enzyme (NEB)	0.5 µL
cDNA template	0.5 µL
Total:	50 µL

Table 3.6 The HU-PCR programmes

98°C	2 min	
98°C	10 s	35 cycles
66.5°C	20 s	
72°C	60 s	
72°C	10 min	
10°C	~	

3.3.9.2 PCR products purification and cloning protocol

Depending on the purpose of the PCR reaction, PCR products were purified by two different methods. If the PCR template was cosmid, plasmid or another pure source, a single band would be found on DNA gel electrophoresis after HU-PCR reaction under the TAE buffer for 60 min at 100 V. In this case, DNA gel cutting and purification was as normal. If the PCR

template was cDNA generated from tissue or cell samples, lots of differently sized PCR bands would be yielded after HU-PCR reaction. In this case, if all the possible alternative splicing transcripts would be captured (while ignoring the genomic BG sequence contamination), PCR products should be run by DNA gel electrophoresis in TAE buffer for only 20 min at 100 V. Under the UV lamp, it could be seen that many DNA bands were not completely separated, and the gel covering sizes ranging from 1500 bp to 3000 bp were cut and purified for cloning.

There were two different methods for handling cloning as well. Depending on the cloning scale, if a large quantity of colonies needed to be picked, the 96 well plate method was applied as described in section 3.2.5. Otherwise, the normal cloning procedure was performed on a small scale.

3.3.9.3 Internal sequencing primers might need be adjusted according to different samples

It was acknowledged through carrying out this project that, although several internal sequencing primers had been designed and successfully applied to get BG cDNA sequenced, new internal sequencing primers might still need to be designed when samples from chickens of new haplotypes were tested due to the polymorphism of BG genes.

3.4 Discussion

In this chapter, a new PCR protocol (HU-PCR) was developed and would replace the previous protocol (H2-PCR) to investigate unknown haemopoietic BG genes from chickens of different haplotypes and to reveal the full length cDNA sequences. Generally speaking, there were three major points showing the superiority of HU-PCR protocol compared to H2-PCR protocol.

Firstly, HU-PCR protocol was easier to use to explore all haemopoietic BG genes. HU primers were designed to target potentially all haemopoietic BG genes no matter whether they had type 1 or type 2 3’ UTRs. On the contrary, H2 primers were only targeting

haemopoietic BG genes with type 2 UTRs, which in theory, another PCR (H1-PCR) has to be done in order to find the haemopoietic BG genes with type 1 3' UTRs. Also, due to the sensitivity of PCR technique, it is very hard to determine the dominantly expressed BG gene from the same sample by performing two different PCRs. Therefore, HU-PCR is superior in evaluating dominantly expressed BG genes than the combination of H1-PCR and H2-PCR.

Secondly, HU primers were proved to be more efficient in amplifying BG cDNAs. Though only two new BG genes were found from duodenum sample of B15 haplotype chicken, many more haemopoietic BG genes plus a tissue BG gene were found from a mixture of duodenum and spleen sample of B12 haplotype using HU primers. Although it could be argued that there is copy number variation between different haplotypes, when the same samples were examined, H2 primers didn't work on sorted T cell or B cells from B19 and B21 haplotypes, while HU primers found several BG genes using the same cDNA templates that were used in H2 PCRs.

Thirdly, HU-PCR protocol provided a good method to investigate alternative splicing of BG cDNAs. Before the establishment of HU-PCR method, there was no systematic study about alternative splicing of any BG gene from BG region, though the phenomena had been observed and was suggested to be exon read through (Kaufman *et al.*, 1990). Using H2 primers, though 20 new BG genes were found by Dr. Chattaway from 4 different tissues (spleen, thymus, bursa and duodenum) of five different haplotypes (B2, B4, B12, B15 and B21), it was not clear whether these new genes contained any alternatively spliced isoform (Chattaway, 2013).

Considering the three aspects discussed, HU-PCR protocol was applied to my whole PhD project whenever new haemopoietic BG genes needed to be investigated. However, it should be noted that another set of universal primers, so-called TU primers, still needs to be designed in order to get all tissue BG genes in the future.

Chapter 4

Functional alleles of chicken BG genes in peripheral T cells

4.1 Introduction

BG genes have been well known for their high polymorphism. Except the two singleton genes, a nearly monomorphic BG0 on chromosome 2 and a polymorphic BG1 in BF-BL region on chromosome 16, all the other known BG genes are located head-to-tail in the BG region on chromosome 16, in which the BG genes in BG region can be easily recombined (Salomonsen *et al.*, 2014; Chattaway *et al.*, 2016). BG genes in BG region have copy number variation, and so far it is unknown which BG genes are orthologous alleles (except for the two singleton genes). Previous research on the B12 haplotype has shown that each BG gene has a striking tissue distribution, so we believe the most dominantly expressed BG genes in a particular cell type from different haplotypes should function the same and that these dominantly expressed BG genes could be treated as 'functional alleles'. By comparing these 'functional alleles' and understanding their variation in sequences, it would give us a better understanding about the polymorphism, whether it is selected, which region of the gene is conserved for function, etc.

BG genes have wide tissue distribution, but their functions still remain unknown. The emerging research on BG homologous molecules [based on extracellular region, the Ig like V domain in human (butyrophilin, BTN) and mouse (skin T cell, SKINT)] have shown strong evidence that BTN and SKINT are involved in γδ T cell regulation (Boyden *et al.*, 2008; Smith *et al.*, 2010; Barbee *et al.*, 2011; Bas *et al.*, 2011; Abeler-Dörner *et al.*, 2012; Rhodes *et al.*, 2015; Rhodes *et al.*, 2016; Salim *et al.*, 2017). BG genes have many similar features as BTN, but are quite different in their cytoplasmic tail, where most BTN molecules have so-called B30.2 domain, while BG genes are famous for 21 nucleotide exons encoding seven amino acid (heptad) repeats to form coiled-coil. It has been noticed that there are different transcripts of BG1 gene from different haplotypes, mainly due to the differences in cytoplasmic tails. It is critically important to understand whether the different transcripts of a BG gene exist within the same haplotype, as it would help us to understand how BG gene form dimers.

Therefore, in this chapter we are aimed to use the universal haemopoietic BG primers, HU primers, to examine BG cDNA sequences from purified T cells of four different chicken lines, and to answer the questions listed as follows.

1. How many BG genes can we find in these four chicken lines?

2. What are the phylogenetic relationships of these genes?

3. Are BG genes conserved between different chicken lines?

4. Comparing to these 'functional alleles', which regions of the BG genes are conversed and which are polymorphic? Is the polymorphism selected?

4.2 Materials and methods

4.2.1 Chicken lines, haplotypes and samples

Four genetic lines of White Leghorn chickens were maintained under specific pathogen-free (SPF) condition at the Institute for Animal Health (IAH): line N, line P2a, line 15I and line 6_1, with the MHC haplotypes of B21, B19, B15 and B2 respectively. The samples used for RNA generation were sorted peripheral T cells from each of the four chicken lines above: the RNA from line N and line P2a by MACS cell sorting (Miltenyi Biotec) was provided by our collaborator Ms Karen Staines; and the RNA from line 15I and line 6_1 by FACS sorting was prepared by myself with Dr. Michaela Fakiola's help during the sorting procedure (FACS was performed by Mr. Nigel Miller in the Department of Pathology). The sorting quality and cDNA integrity were confirmed by PCR amplification of housekeeping gene GAPDH.

4.2.2 RNA isolation, cDNA synthesis and PCR amplification

Roughly $1x10^6$ sorted T cells from each chicken line were extracted for total RNA following the manufacturer's protocol for the NucleoSpin RNA II RNA extraction kit (Machery-Nagel). First strand cDNA was produced from 5-10 ng RNA following the manufacturer's protocol for the Maxima H Minus First Strand cDNA Synthesis Kit (ThermoFisher). Briefly, the RNA was mixed with oligo-$(dT)_{18}$ primer and dNTP mixtures, heated at 65°C for 5 min, chilled on ice for 3 min, 123 RT buffer and Maxima H Minus Enzyme Mix added, and the reaction mixture incubated at 55°C for 45 min, followed by 85°C for 45 min to inactivate the enzyme.

PCR amplification was carried out using Phusion® Hot Start Flex DNA Polymerase (NEB) in a 50 µl reaction mixture with 0.5 µl (5-10 ng) cDNA, 200 µM total dNTPs (1 µl of 10 mM stock), Phusion buffer (10 µl of 5x stock), 1 U Phusion enzyme (0.5 µl of 200U/ml), 0.25 µM forward primer and 0.25 µM reverse primer (both 1.25 µl of 10 µM stocks) and nucleotide-free water (35.5 µl), and with reaction conditions of 2 minute at 98°C, 35 cycles of 98°C 10 s,

66.5°C 20 s and 72°C 60 s, and finally 10 minute at 72°C.

The primers used for PCR amplifications in this chapter were HU primers and SS-TM primers. As described previously in chapter 3, HU primers were designed to amplify the nearly full-length cDNA sequences of BG gene, while the SS-TM primers were designed by Prof. Jim Kaufman to amplify the whole Ig-like V domain and partial signal sequence and partial transmembrane region (Salomonsen *et al.*, 2014). The oligo sequences of HU and SS-TM primers are listed in appendix B.

4.2.3 Cloning and sequencing

The cloning protocol has been described in detail in section 3.3.10.2 (chapter 3). The DNA bands ranging from 1500 to 3000 bp by 1% agarose gel electrophoresis after HU-PCR reaction were purified and cloned into pJET vector (CloneJET PCR cloning kit, ThermoFisher), and 92-96 colonies were picked for colony PCR using either HU primers or SS-TM primers. Positive clones were grown as bacterial cultures for miniprep (PureLink Quick Plasmid Minipre Kit, Invitrogen) and sent for dideoxy chain termination sequencing (DNA Sequencing Facility, Department of Biochemistry, University of Cambridge). Sequencing primers used in this chapter were T7, pJETR, UC699, UC700, UC701 and UC703 with the detailed oligo sequences listed in appendix B.

4.2.4 Sequence analysis

Most sequence analyses, including sequencing data organization, annotation, alignment, comparison and phylogenetic study, were introduced from section 3.2.5 to 3.2.7 in chapter 3, while three extra operations performed in this chapter are described as below.

First, the output of sequence alignment was edited in BioEdit [BioEdit Sequence Alignment Editor (http://www.megasoftware.net/)]. Sequence alignments, no matter whether generated by CLC DNA Workbench or MEGA7, were imported into BioEdit, then exported as a 'rich text with current shaded view setting', opened in Word (Microsoft) and modified manually by adding annotations.

Second, the helical wheel analyses for the cytoplasmic tails were done using online tool DrawCoil10 (http://www.grigoryanlab.org/drawcoil/), and the results were taken into account when manually drawing the cartoon for coiled-coil illustration.

Third, the model of Ig-like V domain of BG8-B12 was built using SWISS-MODEL online tool (https://swissmodel.expasy.org/interactive) based on the template with pdb ID of 3csp.1. The template was a crystal structure of human butyrophilin subfamily 3 member A3 (BTN3A3), which shares 43.36% amino acid sequence identity with our model in the modeling region, indicating the structure built was highly reliable. The model was then visualized, rendered and modified in PyMOL (http://pymol.org/2/) installed locally.

All the other figures in this chapters were designed and manipulated in Word or PowerPoint (Microsoft).

4.3 Results

4.3.1 HU PCR results and using exon 2 to distinguish different BG genes

Total RNA isolated from peripheral T cells of four different chicken lines was reverse-transcribed into cDNA separately using oligo-dT primer; cDNAs were amplified by two independent PCRs using HU primers; the amplicons were cloned and sequenced. Many BG sequences were found with different lengths but all had the organization of typical mRNA with nearly full 5'UTRs and 3'UTRs. A third independent PCR was performed for line 6_1 using SS-TM primers in order to evaluate the results. For each PCR reaction, all the fully sequenced clones were compared, and the ones with exactly the same nucleotide sequence were grouped. One representative clone was selected to represent all the other clones in the group, and a number was added into the representative's name to indicate how many clones in total showed such sequence.

4.3.1.1 The complexity of sequences

At first glance of all the complicated sequences, there was no clue which sequence belonged to which BG gene. To make sequence alignments easier, clones from all PCRs of a same chicken line were compared first. The ones with exactly the same nucleotide sequence were grouped and the sequence was renamed with all representative clone names from each PCR organized following the order of first PCR, second PCR and third PCR, and a number followed to indicate how many clones in total showed such sequence. However, after the reorganization there were still about hundred different sequences from all four chicken lines with variation going through the whole cDNA structures, bringing huge difficulties for

further analysis.

4.3.1.2 Exon 2 serves well to distinguish different BG genes

After many failures of trying to identify how many BG genes were there, a method of comparing each sequence based on exon 2, containing the last two nucleotides of the last codon of the signal sequence and the whole Ig-V domain, successfully categorized all these sequences into 16 BG genes, revealing that all the variation described above were actually due to alternative splicing events in the cytoplasmic tail regions. Thus, exon 2 sequences were used to distinguish different BG genes, which proved to serve very well for all the BG sequences acquired so far.

4.3.1.3 BG0 and BG1 were detected

During each PCR in this project, either BG0 or BG1 or both were observed in our samples which was different to the result of our previous tissue distribution work on B12 haplotype using SS-TM primers where neither BG0 nor BG1 was detected in T cells (Salomonsen *et al.*, 2014). SS-TM primers were considered as 'universal primers', which later was found not to work for BG0 or BG2; therefore specific primers were designed to retest these samples from which BG0 was found in both purified T cells and B cells. In our project objectives, BG0 and BG1 expression study was not part of our original plan; therefore, when the HU primer pair was designed, neither BG0 nor BG1 was considered to be amplified. However, against our expectations, BG0 and BG1 cDNAs were found in our samples also with alternative splicing seen in their cytoplasmic tails, while in the cytoplasmic tail region of BG1 cDNA sequence, it has been known already that there is an ITIM motif. Further, BG0 and BG1 were the two singleton BG genes found in the same genomic locations of every haplotype of chicken, each with only one copy. Considering the special and important roles they might take in T cells and B cells, it would be worth exploring their cDNA transcripts in detail.

4.3.2 Nomenclature of BG sequences in this chapter

Until now, there has been no unified nomenclature for all BG genes found, except two singleton BG genes, BG0 on chromosome 2 and BG1 in BF/BL region on chromosome 16. The naming of the 14 BG genes found in B12 haplotype was the first and only time that complete BG genes from one haplotype were systematically annotated, compared and

analyzed. However, due to the polymorphism and copy number variation of BG genes in the BG region, as well as the lack of genomic typing for BGs from other haplotypes, it is difficult to give any BG gene from BG region a fixed name at this moment. In order to make our data easily understood, each gene identified in this project was given a name with the abbreviated line name, 'T' for T cells, 'BG' and the letter 'a' representing the most frequently detected clone from the most frequently detected exon 2 sequence (and 'b' being the second most frequently detected exon 2 sequence, and so forth). For example, NTBGa is the most frequently detected gene in T cells from line N. In addition, each transcript described in this chapter was given its gene name (which gene it belonged to), followed by a dash and then a number with '1' being the most frequently detected transcript ('2' for the second mostly frequently detected transcript, and so forth), and sometimes with numbers in parentheses indicating the colony number (how many clones showed the same sequence), as well as the numbers of PCRs (how many independent PCR reaction showed this sequence).

4.3.3 Example of how many BG genes were found in line N and the nomenclature

Two independent HU-PCRs were performed on sorted T cells of line N (B21 haplotype); the amplicons from each PCR were cloned in separate ligations. In each cloning, 94 clones were picked to run colony PCR, and only the clones confirmed to carry BG fragment insertions were grown in bacterial cultures for miniprep and sequencing. If, for any reason, the experiment failed or the results were in doubt, the whole procedure was repeated until clones with BG cDNA transcripts were obtained and fully sequenced. All the clones were named with abbreviation of 'N' for line N, 'T' for T cell, followed by either '1' or '2' standing for 1st or 2nd PCR reaction, and the colony ID number. For example, 'NT108' was the name for number 08 colony which was generated from 1st PCR reaction of T cells from line N.

All the clones from the same PCR reaction were compared first, and the ones with exactly the same nucleotide sequence were grouped. One representative clone was selected to represent all the other clones in the group, and a number was added into the representative's name to indicate how many clones in total showed such sequence. For example, NT106(x13) was the representative for another 12 clones which had the same nucleotide sequence as NT106. Then all the representative sequences were aligned and categorized into different BG gene file folders according to their exon 2 sequences. After such procedures were done for both of the two independent PCRs, all the representative sequences were compared and the same sequences were grouped. As before, a new name was given and colony numbers were added,

however, the difference was, in order to understand this sequence came from both two independent PCRs, both the representative clones names were kept in the new name. For example, NT105201(x35) was a sequence which appeared in both two independent PCR amplifications, with clone 105 and clone 201 from the first PCR and second PCR reaction respectively sharing the same nucleotide sequence, and altogether there were 35 clones carrying this sequence.

After all clones were categorized into different BG file folders based on their exon 2 sequences, the clone numbers were counted. The file folder with most frequently detected exon 2 sequence was named NTBGa, meaning all the cDNA transcripts found within this particular exon 2 sequence belongs to the gene NTBGa, the second most frequently detected gene was named NTBGb, and so the forth. As there are five different BG genes (based on exon 2 sequences), they were named with NTBGa, NTBGb, NTBGc, NTBGd and NTBGe, respectively. Under each gene, the most frequently detected cDNA transcript was named with the number '1', followed by '2', and so forth. For example, NTBGa-1(x35) was the most frequently detected cDNA sequence of gene NTBGa; in order to understand whether this transcript was observed in only one PCR reaction or in both two PCR reaction, as well as how many clones in total belongs to NTBGa, the total colony numbers and PCRs were added in some circumstances. For example, from the name of NTBGa-1(35/68, 2) it could be interpreted that this was the most frequently detected cDNA transcript from gene NTBGa which has 35 clones out of 68 in total showing this sequence, and it was found in both independent PCRs.

To summarize the names used in this chapter, some examples were listed as below.

NTBGa: BG gene 'a' (NT**BGa**) from sorted T cells (N**T**BGa) of line N (**N**TBGa).

NTBGa(68/76): BG gene 'a' (NT**BGa**) from sorted T cells (N**T**BGa) of line N (**N**TBGa), and 68 colonies [NTBGa(**68**/76)] showed NTBGa in that PCR reaction with total colony number of 76 [NTBGa(68/**76**)].

NTBGa-1: Isoform 1 (NTBGa-**1**) of BG gene 'a' (NT**BGa**-1) from sorted T cells (N**T**BGa-1) of line N (**N**TBGa-1).

NTBGa-1(35/68, 2): Isoform 1 [NTBGa-**1**(35/68, 2)] of BG gene 'a' [NT**BGa**-1(35/68, 2)] from sorted T cells [N**T**BGa-1(35/68, 2)] of line N [**N**TBGa-1(35/68, 2)], and 35 colonies

[NTBGa-1(**35**/68, 2)] showed NTBGa-1 in total of 68 clones [NTBGa -1(35/**68**, 2)] showing NTBGa, and this sequence has been detected in two independent PCRs [NTBGa-1(35/68, **2**)].

4.3.4 Sixteen BG genes were found from T cells of four chicken lines

In total, there were 16 different BG genes based on exon 2 sequences found from sorted peripheral T cells of four chicken lines, line N (B21), line P2a (B19), line 15I (B15) and line 6_1 (B2) with 15 found by HU primers and the other one found by SS-TM primers. In each line, there were 3-5 genes expressed, most with lots of alternative splicing transcripts in the cytoplasmic tail regions, but no genes were shared among the four chicken lines.

4.3.4.1 Five BG genes were found in line N (B21)

In line N (B21), five BG genes (NTBGa, NTBGb, NTBGc, NTBGd and NTBGe) were expressed in T cells. As shown in table 4.1, only NTBGa was found in both PCRs with 16 different nearly full-length cDNA transcripts, most due to alternative splicing in the cytoplasmic tails. Some of these transcripts appeared in both PCRs while most of them were found only in one PCR and some even with only one clone. Complicated as it was, it was hard to rule out many of these sequences with only one clone (except the ones with obvious PCR errors), because some of them were proved to exist in later PCR amplifications using B cell cDNA generated from the same blood sample of line N (data not shown in this chapter). Therefore, though it might be argued that PCR-based mutations in NTBGa-1 were likely to have generated NTBGc (position 242, 542 and insertion at position 906), and in NTBGa-4 to generate NTBGb (position 173 and 1660) (Appendix D), NTBGb and NTBGc were kept as individual genes as the mutations took place in exon 2 region. However, the sequence information revealed by this study could be used for future research through more unbiased approaches such as RNAseq, genomic sequencing or BG gene typing, which eventually could answer this question.

Table 4.1 Summary of all BG cDNA transcripts found in line N (B21)

Line N (B21)	Representative Clones	New Names
NTBGa(68/76)	seq1-1(x35)NT105201	NTBGa-1(35, 2)
	seq1-2(x8)NT234	NTBGa-2(8, 1)
	seq1-3(x6)NT127235	NTBGa-3(6, B)
	seq1-4(x5)NT125	NTBGa-4(5, B)
	seq1-5(x2)NT111	NTBGa-5(2, 1)
	seq1-6(x2)NT104	NTBGa-6(2, 1)
	seq1-7(x1)NT101	NTBGa-7(1, 1)
	seq1-8(x1)NT132	NTBGa-8(1, 1)
	seq1-9(x1)NT108	NTBGa-9(1, 1)
	seq1-10(x1)NT208	NTBGa-10(1, 1)
	seq1-11(x1)NT210	NTBGa-11(1, 1)
	seq1-12(x1)NT212	NTBGa-12(1, 1)
	seq1-13(x1)NT211	NTBGa-13(1, 1)
	seq1-14(x1)NT221	NTBGa-14(1, 1)
	seq1-15(x1)NT239	NTBGa-15(1, 1)
	seq1-16(x1)NT231	NTBGa-16(1, B)
NTBGb(4/76)	seq2(x4)NT237	NTBGb(4, 1)
NTBGc(2/76)	seq3(x2)NT110	NTBGc(2, 1)
NTBGd(1/76)	seq4(x1)NT138	NTBGd(1, 1)
NTBGe(1/76)	seq6(x1)NT217	NTBGe(1, 1)

Note: Left column: the summary of the BG genes found in line N with the numbers of how many clones were found for this gene out of a total of how many clones were found from line N; the left column being coloured indicated the gene was found in at least two independent PCRs, otherwise the column keeps white background. Middle column: the original record of representative clones that show a particular BG cDNA sequence with the numbers showing how many clones were found with such sequence. Right column: the final names representing the particular BG cDNA sequence on the left. The final names follow the convention: 'N' for line N, 'T' for T cells, 'BG' and a letter representing the exon 2 sequence, a dash and then a number representing the alternative splicing variant with '1' being the most frequently detected clone (and '2' the second most frequently detected clone, and so forth); the first number in parentheses indicates the number of clones found for a particular exon 2 sequence, and the second number or letter indicates from how many independent PCRs these clones were found with '1' for one PCR, '2' for two PCRs, 'B' for one PCR in B cells (chapter 5) and one PCR in T cells.

4.3.4.2 Three BG genes were found in line P2a (B19)

Three BG genes, PTBGa, PTBGb and PTBGc were found from line P2a (B19) in both PCRs (Table 4.2). The most dominantly expressed cDNA transcripts under P2aTBGa and P2aTBGb were found in both PCRs. However, all six cDNA transcripts of P2aTBGc were found with only one clone each, except one cDNA transcript was found in later PCR experiment using B cell cDNA generated from the same blood sample of line P2a.

Table 4.2 Summary of all BG cDNA transcripts found in line P2a (B19)

Line P2a (B19)	Representative Clones	New Names
P2aTBGa(14/26)	seq1-1(x10)PT102221	P2aTBGa-1(10, 2)
	seq1-2(x3)PT122	P2aTBGa-2(3, 1)
	seq1-3(x1)PT202	P2aTBGa-3(1,1)
P2aTBGb(6/26)	seq2-1(x2)PT104218	P2aTBGb-1(2, 2)
	seq2-2(x2)PT109	P2aTBGb-2(2, 1)
	seq2-3(x1)PT218	P2aTBGb-3(1, 1)
	seq2-4(x1)PT125	P2aTBGb-4(1, 1)
P2aTBGc(6/26)	seq3-1(x1)PT117	P2aTBGc-1(1, B)
	seq3-2(x1)PT107	P2aTBGc-2(1, 1)
	seq3-3(x1)PT103	P2aTBGc-3(1, 1)
	seq3-4(x1)PT223	P2aTBGc-4(1, 1)
	seq3-5(x1)PT211	P2aTBGc-5(1, 1)
	seq3-6(x1)PT228	P2aTBGc-6(1, B)

Note: same as legend to table 4.1.

4.3.4.3 Three BG genes were found in line 15I (B15)

Three BG genes were obtained in line 15I (B15) with 15iTBGa and 15iTBGb found in both PCRs, and 15iTBGc only found in one PCR reaction (Table 4.3). However, 15iTBGc was also found in later PCR experiment using B cell cDNA generated from the same blood sample of line 15I. Likewise, many of the cDNA transcripts that only showed up in one PCR here also appeared in the B cell project.

Table 4.3 Summary of all BG cDNA transcripts found in line 15I (B15)

Line 15I (B15)	Representative Clones	New Names
15iTBGa(24/39)	seq1-1(x6)L15T239	15iTBGa-1(6, B)
	seq1-2(x4)L15T127208	15iTBGa-2(4, 2)
	seq1-3(x3)L15T118216	15iTBGa-3(3, 2)
	seq1-4(x3)L15T102	15iTBGa-4(3, B)
	seq1-5(x2)L15T117	15iTBGa-5(2, B)
	seq1-6(x2)L15T226	15iTBGa-6(2, 1)
	seq1-7(x1)L15T140	15iTBGa-7(1, B)
	seq1-8(x1)L15T105	15iTBGa-8(1, 1)
	seq1-9(x1)L15T218	15iTBGa-9(1, 1)
	seq1-10(x1)L15T134	15iTBGa-10(1, 1)
15iTBGb(11/39)	seq2-1(x4)L15T126207	15iTBGb-1(4, 2)
	seq2-2(x2)L15T128215	15iTBGb-2(2, 2)
	seq2-3(x1)L15T123	15iTBGb-3(1, 1)
	seq2-4(x1)L15T209	15iTBGb-4(1, 1)
	seq2-5(x1)L15T219	15iTBGb-5(1, B)
	seq2-6(x1)L15T227	15iTBGb-6(1, 1)
	seq2-7(x1)L15T237	15iTBGb-7(1, 1)
15iTBGc(4/39)	seq3-1(x3)L15T201	15iTBGc-1(3, 1)
	seq3-2(x1)L15T210	15iTBGc-2(1, 1)
	seq3-2(x1)L15T210	15iTBGc-2(1, 1)

Note: same as legend to table 4.1.

4.3.4.4 Five BG genes were found in line 6_1 (B2)

Five BG genes were obtained from line 6_1 (B2) with 6TBGa, 6TBGb, 6TBGc, 6TBGd and 6TBGe in three PCRs, with two PCRs using HU primers and the other one using SS-TM primers (Table 4.4). There was no gene really overwhelmingly dominant as found previously in B12 haplotype or the other haplotypes in this project. It is worth clarifying that the clone numbers counted in table 4.4 includes the clones found in SS-TM PCR, from which the sequences of these clones only contain partial signal sequence, the whole Ig-V domain and partial transmembrane region. The 6TBGd gene was only found in SS-TM PCR; thus some analyses later regarding the full length cDNA sequence of 6TBGd were lacking.

It is very interesting that the nearly full-length conceptual transcripts (that is, exons without introns) from four BG genes found in line 6_1 (B2) are also found in the B12 haplotype, though none of the 16 BG genes found from the four chicken lines in this project was identical to each other. As shown in appendix E, nucleotide sequence alignments between the most dominantly expressed cDNA transcripts of genes from line 6_1 and their counterparts from B12 listed from (A) to (E) clearly reflected that each pair consisted of the same genes with very few differences which were caused by PCR errors or alternative splicing or intron read-through in the cytoplasmic tails. For example, there was only one nucleotide difference in the cytoplasmic tail region of 6TBGb-1 and BG7-B12 pair, which might be caused by PCR error. In the rest of the pairs there were alternative splicing or intron read-through in the cytoplasmic tail region of line 6_1 BG genes compared to B12 BGs. However, it should be noted that the B12 BG cDNA sequences were derived from genomic cosmid sequencing results, so they might not represent the expressed cDNA transcripts.

In fact, the identical BG genes found between line 6_1 (B2) and line CB (B12) support previous findings. Dr. Simonsen had noted the serological identity of B2 and B12 molecules on erythrocytes long time ago (Simonsen *et al.*, 1982). Dr. Miller made BG antigen 2-D gels of many different haplotypes, and by comparing the 2-D gel patterns of each haplotype, she found B2 and B12 haplotype were quite similar (Miller *et al.*, 1984). Here we found four BG genes in line 6_1 (B2) were the same as the ones in B12 haplotype, so it is not surprising that the BG antigens from the two different haplotypes look very similar.

However, the mystery is that the dominantly expressed BG gene in B12 haplotype was BG9 with a few BG12 clones, while these two genes did not show up in line 6_1 at all, and instead,

the dominantly expressed BG gene was BG13 which was quite different from BG9 or BG12. Detailed analysis is carried out in section 4.3.6.

Table 4.4 Summary of all BG cDNA transcripts found in line 6_1 (B2)

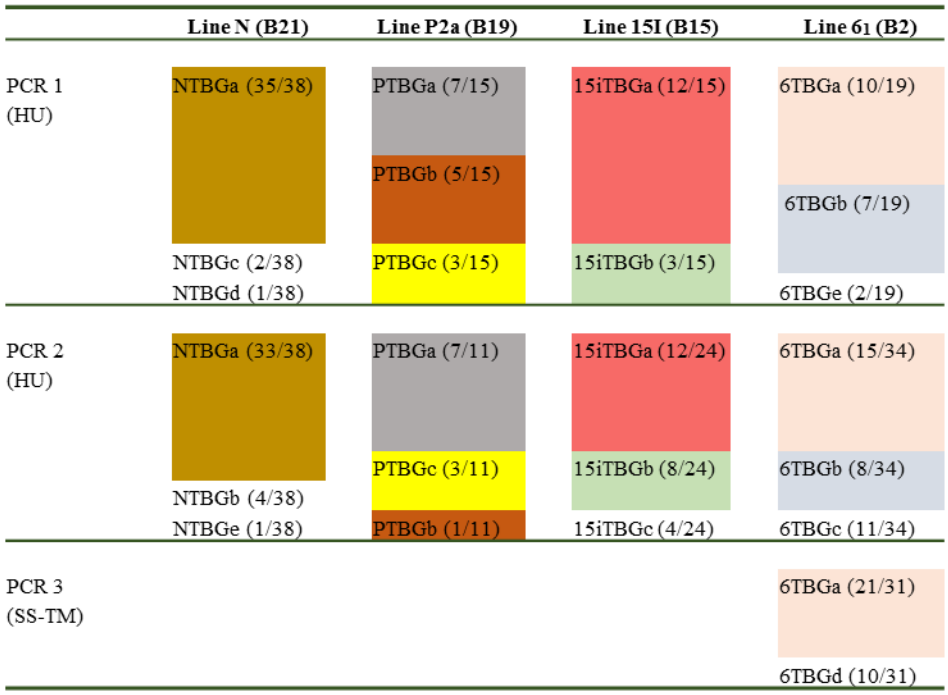
Line 6_1 (B2)	Representative Clones	New Names
6TBGa(46/84)	seq1(x10)L6T108	6TBGa-1(10, 1)
	seq1-1(x9)L6T211	6TBGa-2(9, 1)
	seq1-2(x3)L6T224	6TBGa-3(3, 1)
	seq1-3(x2)L6T227	6TBGa-4(2, 1)
	seq1-4(x1)L6T222	6TBGa-5(1, 1)
	seq1(x21)L6T337	6TBGa-6(21, V)
6TBGb(15/84)	seq2(x7)L6T118	6TBGb-1(7, B)
	seq2-1(x6)L6T212	6TBGb-2(6, 1)
	seq2-2(x2)L6T221	6TBGb-3(2, 1)
6TBGc(11/84)	seq3-1(x6)L6T230	6TBGc-1(6, B)
	seq3-2(x3)L6T206	6TBGc-2(3, 1)
	seq3-3(x2)L6T228	6TBGc-3(2, 1)
6TBGd(10/84)	seq1(x10)L6T321	6TBGd(10, V)
6TBGe(2/84)	seq3(x2)L6T119	6TBGc-3(2, 1)

Note: same as legend to table 4.1 except 'V' for one PCR but only tested for the SS-TM region using SS-TM primers.

4.3.4.5 Summary: 16 BG genes were found in four different chicken lines with most genes having alternative splicing in cytoplasmic tails

As shown in figure 4.1, sixteen different BG genes (based on exon 2 sequences) were detected from peripheral T cells of four different chicken lines: line N (B21), line P2a (B19), line 15I (B15) and 6_1 (B2). For each line, two independent PCRs were performed using HU primers, and one extra PCR reaction was performed on line 6_1 using SS-TM primers. From each line, one dominant and several subdominant BG genes were found, however only line N showed one overwhelming dominantly expressed BG gene like previous results for line CB (B12). None of the 16 genes were found in more than one line except four BG genes from line 6_1 were the same as the genes well characterized in line CB (B12); however, the mystery is that the genes expressed in T cells from line CB did not appear in T cells from line 6_1 at all.

Another interesting finding was that alternatively spliced cytoplasmic tails were observed in most of these genes' cDNA transcripts. After aligning and comparing all these nearly full-length transcripts, ruling out the obvious PCR chimeras, alignments of 57 BG cDNA transcripts from all the 16 BG genes were made in appendix D, from which the detailed sequence information as well as the clone information (how many clones showed this sequence in how many PCRs) could be referred to.



Line CB (B12)	Line N (B21)	Line P2a (B19)	Line 15I (B15)	Line 61 (B2)
BG9 (25/28)	NTBGa (68/76)	P2aTBGa (14/26)	15iTBGa (24/39)	6TBGa (46/84)
BG12 (3/28)	NTBGb (4/76)	P2aTBGb (6/26)	15iTBGb (11/39)	6TBGb (15/84)
	NTBGc (2/76)	P2aTBGc (6/26)	15iTBGc (4/39)	6TBGc (11/84)
	NTBGd (1/76)			6TBGd (10/84)
	NTBGe (1/76)			6TBGe (2/84)

Figure 4.1 Overall results for the number of BG genes amplified from T cells of four chicken lines. BG genes (based on exon 2 sequences) were amplified from cDNA preparations derived from peripheral T cells isolated from various chicken lines (with different B haplotypes): line N (B21), line P2a (B19), line 15I (B15) and line 6_1 (B2). Top panel, independent amplifications from the four chicken lines; HU, haemopoietic forward and 'universal' reverse primers to give nearly full-length sequences; SS-TM, signal sequence forward and transmembrane reverse primers to give SS, extracellular Ig-V domain and TM regions. Different colours indicate different exon 2 sequences, except those sequences that are only found in one PCR reaction. Names follow the convention: abbreviated line name, 'T' for T cells, 'BG', and a letter representing the exon 2 sequence; numbers in parentheses indicate the number of clones found for a particular exon 2 sequence out of the total number for the particular PCR reaction. Bottom panel, the total results for four chicken lines from this project and for the CB line (B12) from Salomonsen *et al*., 2014.

4.3.5 BG genes are highly polymorphic and phylogenetic studies show different relationships of BG genes in different regions

Regardless of the alternative splicing in cytoplasmic tails found in the BG cDNA transcripts, the high polymorphism of BG genes was proved by many nucleotide differences through-out each region of the cDNA sequences among the 16 BG genes found in this project (Appendix D). In order to understand the phylogenetic relationships of these BG genes at genetic level, the conceptual cDNA sequences (that is exons without any intron) of each gene (Appendix F) were applied for phylogenetic study using Neighbour Joining method with bootstrap (1000). Overall, by comparing to all 14 BG genes from B12 haplotype using the full-length cDNA sequences, all 15 genes (6TBGd was not included because of the lack of full cDNA sequence) have haemopoitic 5' UTRs. Twelve genes belong to the BG8-9-12-13 clade, and the other three belong to the BG5-7-11 clade (Figure 4.2). However, the phylogenetic trees built on separate regions showed quite different stories (Figure 4.3).

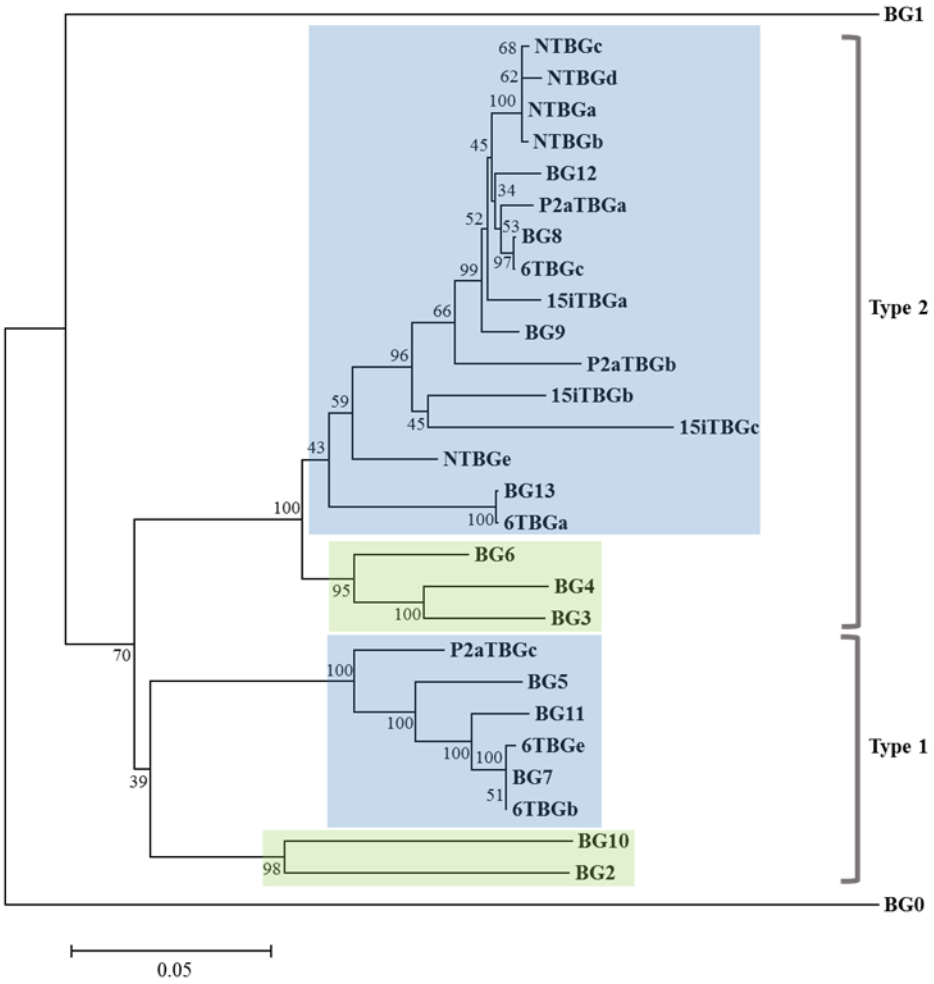


Figure 4.2 Phylogenetic tree built on full-length conceptual transcripts of 15 BG genes found in T cells of four chicken lines and all 14 BG genes from B12 haplotype. Names of the transcripts follow the convention: abbreviated line name, 'T' for T cells, 'BG' and the letter 'a' representing a particular exon 2 sequence. Names of the genes follow the convention 'BG' and the number of the gene locus from the B12 haplotype. Indicated by colour are those clades with 5' ends of haemopoietic (blue) and tissue (green), and by brackets for 3' ends of type 1 and type 2. Branch lengths are scaled by genetic distance, and percentage bootstrap values are indicated at the nodes. Note: 6TBGd is not present in this tree since it was only detected by the SS–TM amplification.

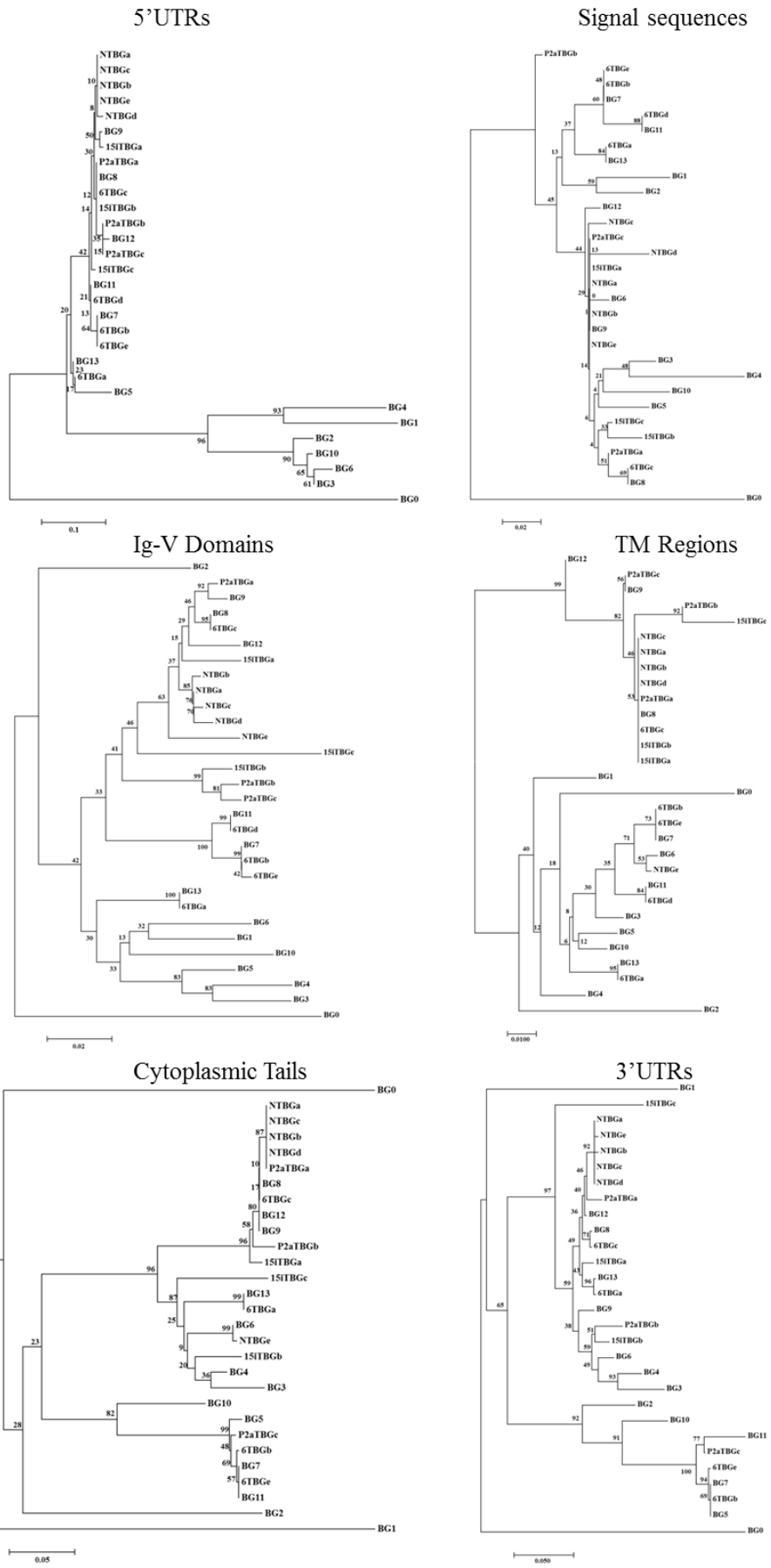


Figure legend is shown on the page following the figure

Figure 4.3 Phylogenetic trees of nucleotide sequences for different regions of the full-length conceptual transcripts for all 16 BG genes found in T cells from four chicken lines and 14 BG genes from B12 haplotype. Phylogenetic trees include 5' UTR from exon 1, signal sequence from exon 1, exon 2 (two nucleotides from the signal sequence, and then the nucleotides encoding the Ig-V domain), exon 3 (transmembrane region), the exons corresponding to the cytoplasmic tail (excluding any nucleotides in the final exon that encode amino acids), and the final exon (which is exactly the 3' UTR in some sequences, but for which the first nucleotides encode the last amino acids of the cytoplasmic tail in most sequences). Other details are as in the legend to figure 4.2.

4.3.5.1 Phylogenetic tree of 5'UTRs of BG genes falls into two groups

As shown in figure 4.3, 5'UTRs of BG genes fall into two groups and all 15 BG genes found in this project together with the 'Haemopoietic BGs' of B12 haplotype are in the same group, indicating all 15 BG genes are also haemopoietic BGs. According to previous study on B12 BG genes, the phylogenetic tree, built on 5'UTRs of all 14 BG genes, falls into two groups, correlated with the tissue distribution pattern. Therefore one group of genes expressed in haemopoietic cells were called 'Haemopoietic BGs' while the other group of genes expressed in tissues were called 'Tissue BGs'. As all 15 BG genes are found from T cells with the universal primer pair aiming for amplifying haemopoietic BGs, it is reasonable that they fall into the 'Haemopoietic BGs' group.

One major difference between 'Haemopoietic BGs' and 'Tissue BGs' in their 5'UTRs is the large deletion in 'Haemopoietic BGs', which is considered as a true deletion by Salomosen and his colleagues, who also propose that all the 'Haemopoietic BGs' in the B12 haplotype are descended from a single ancestor (Salomonsen *et al.*, 2014). This assumption is further supported by the result of alignment here (Figure 4.4) that all 'Haemopoietic BGs' from four chicken lines also have the same large fragment deletion compared to 'Tissue BGs'.

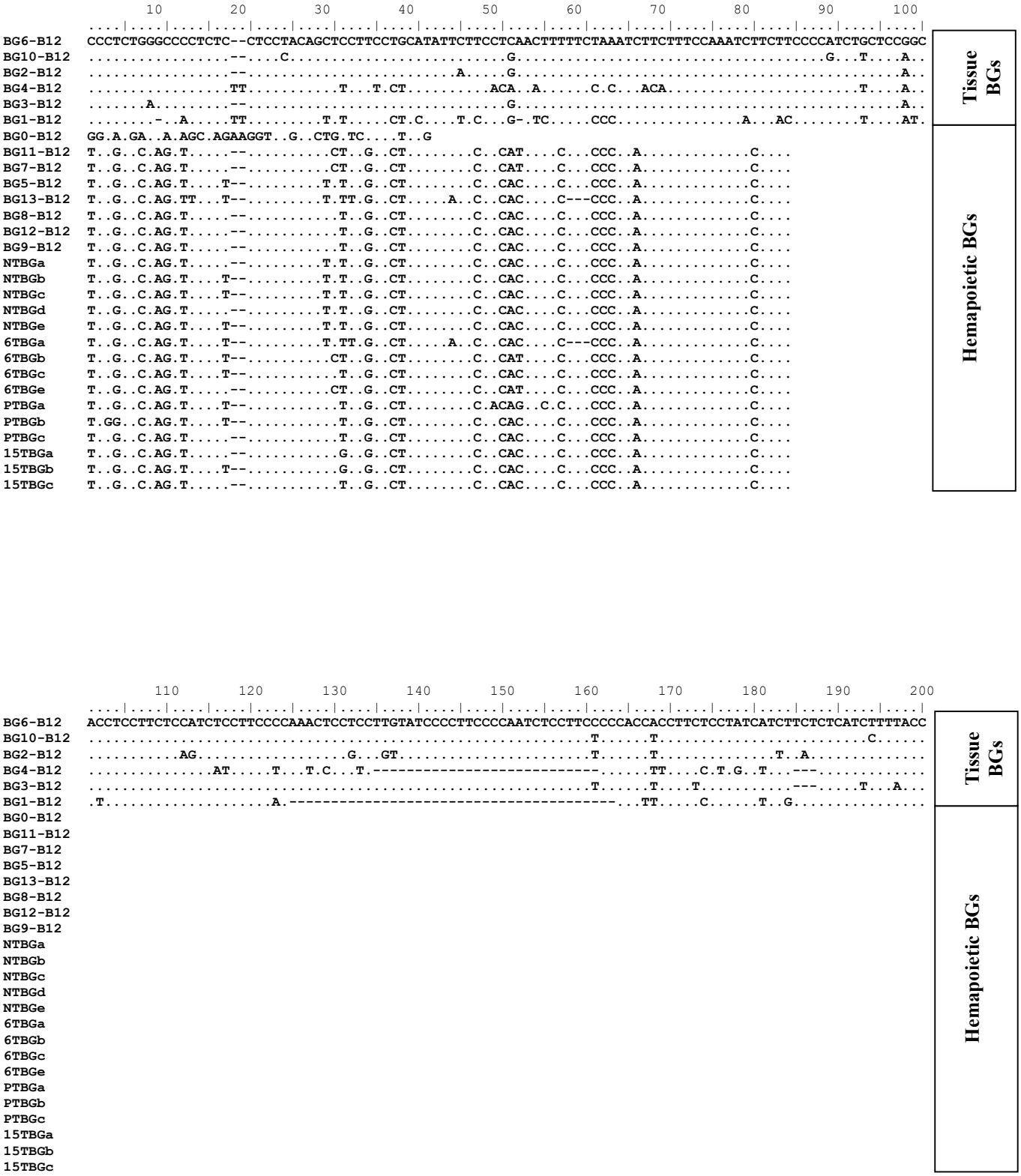
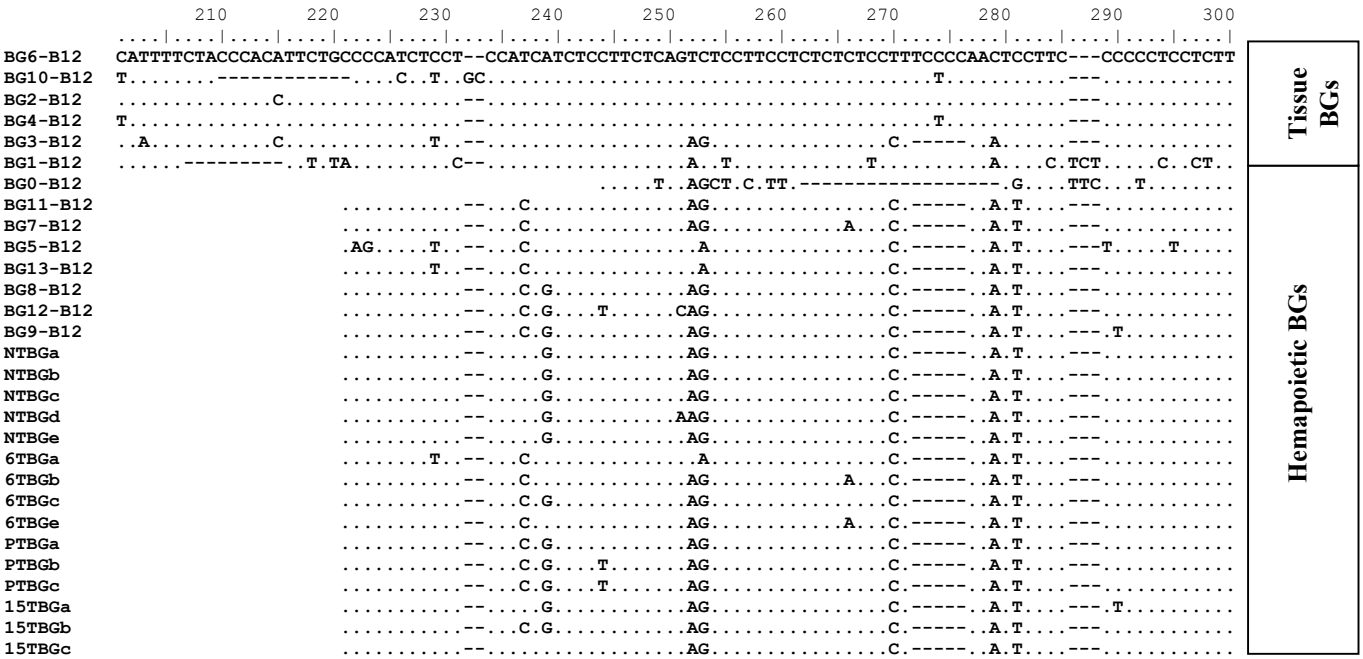


Figure legend is shown on the page following the figure



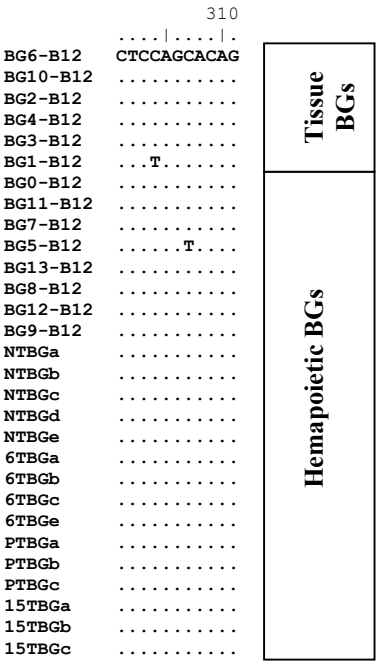


Figure 4.4 The alignment of 5' UTRs between 15 BG genes found in T cells of four chicken lines and all 14 BG genes from B12 haplotype. Compared to the tissue BGs, there is a large gap in the 5' UTR sequences of haemopoietic BGs; the alignments showed that all the 15 BGs found in this project have haemopoietic 5'UTRs.

4.3.5.2 Phylogenetic tree of signal sequences do not fall into clear groups

The signal sequences are too short (99 bp) to produce a reliable tree, and this tree doesn't show any pattern or differentiation to help understand the relationships of all the genes. However the tree is also presented in figure 4.3.

4.3.5.3 Phylogenetic tree of Ig-V domains shows most BG genes from the 16 new genes are clustered within BG8, 9 and 12 group

Although high polymorphism is shown in the Ig-like V domain (Appendix F), the phylogenetic tree built with this region seems less complicated than expected, in which most BG genes found in this project are distributed into the three groups identified already in the V domain dendrogram built on B12 BGs (Salomonsen *et al.*, 2014). In Salomonsen's work, the three groups with very similar V sequences are: BG3, 4 and 5; BG7 and 11; BG8, 9 and 12. Here in our tree (Figure 4.3) which includes 6TBGd as it has complete Ig-V domain sequence, the majority of genes stay in the BG8, 9 and 12 group, three genes belong to BG7 and 11 group, and one gene 6TBGa is with BG13 as they are identical. It is worth mentioning that previous study has showed that BG13 gene together with BG8, BG9 and BG12 are overall closely related and expressed in the same kinds of cells and tissue, and their Ig-V domains differ in two stretches of sequences which might be caused by micro-recombination between BG8-9-12 gene and BG6 (Chattaway *et al.*, 2016). Unlike the dominant expressed genes in the other three chicken lines (line N, line P2a and line 15I) for which their Ig-V domains belong to BG8-9-12 cluster, the dominant expressed gene in line 6_1, 6TBGa, is identical to BG13, which will be examined in detail in section 4.3.6.

4.3.5.4 There are two types of transmembrane regions

The transmembrane regions are too short (105 bp) to reflect a very reliable phylogeny, especially when many genes had virtually identical sequences. However, by looking at the tree, it seems that there are two groups, one group including BG8-9-12 and the other group including all other B12 BG genes. Most genes found in this project fall into the BG8-9-12 group, while all BG genes from line 6_1 (except 6TBGc) together with one gene from line N (NTBGe) go to the other group. Further analysis based on amino acid comparison will be discussed in section 4.3.6.

4.3.5.5 There are type 1, type 2a and type 2b cytoplasmic tails

Previous research based on B12 haplotype found that there were so called two types of cytoplasmic tails, type 1 and type 2, which correlated with the two types of 3'UTR (Chattaway, 2013). From our phylogenetic tree for cytoplasmic tails, it is obvious that within the type 2 sequences, there are two subtypes, type 2a and type 2b. All 14 BG sequences found in our project fall into three types, where P2aTBGc, 6TBGb and 6TBGe are in type 1 which also include BG5, 7, 11 and 10 from B12 haplotype; NTBGe, 15iTBGb, 15iTBGc and 6TBGa are in type 2b which also include BG3, 4, 6 and 13; and the rest of the genes with BG8, 9 and 12 from B12 haplotype constitute the type 2a.

4.3.5.6 Two types of 3'UTRs were identified from phylogenetic tree on 3'UTRs

All BG genes are clustered into two groups in their 3'UTR regions which is consistent with the two types of 3'UTR from previous work (Salomonsen *et al.*, 2014). It also fits well to the concept that these two 3'UTR types were not related to 'Haemopoietic' or 'Tissue' BGs as each of them contain BG genes either with type 1 or type 2 3'UTRs. Also the results show that the HU primer pair can amplify haemopoietic BG genes with either type 1 or type 2 3'UTRs.

4.3.5.7 Summary: different regions illustrate different relationships among all the BG genes analyzed

From the phylogenetic studies based on conceptual cDNA transcripts of the BG genes found in this project ('the new BG genes') as well as the well characterized 14 BG genes from B12 haplotype, there were at least three messages generated. First, all the new BG genes belong to Haemopoietic BGs with either type 1 or type 2 3'UTR, and most of these genes are from the BG8-9-12-13 clade of B12 haplotype, which matches the fact that these genes were found from T cells where in B12 haplotype it was BG8-9-12-13 clade that hold the most genes found from either T or/and B cells. Second, phylogenetic trees made with different regions reflected different stories as the genes clustered in the same group in one region may be clustered into different groups in another region, which may also explain the third point. Third, BG genes look like hybrid genes, which had already been discovered by our lab through exploring BG cDNA transcripts from tissue samples (Chattaway, 2013). As shown in the diagram of figure 4.5, it could be seen that all the BG genes look pretty much like hybrid

genes, some of them were close to each other in one region but set far away from each other in another region, which again provides evidence to support the idea of deletion and recombination during gene expansion and contraction of BG evolution (Salomonsen *et al.*, 2014).

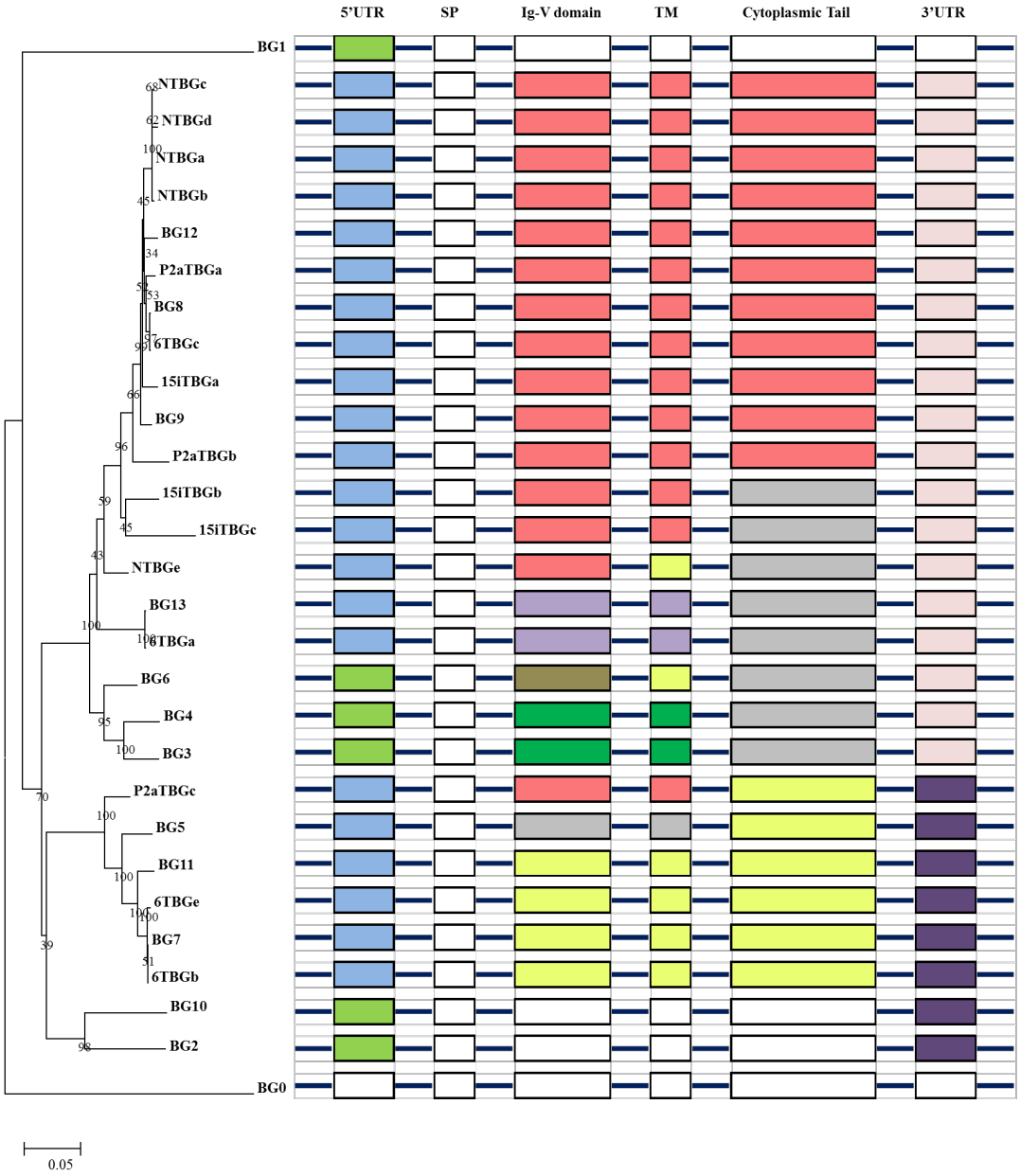


Figure 4.5 Overview of the phylogenetic relationships of 15 BG genes found in T cells from four chicken lines with 14 BG genes from B12 haplotype. Phylogenetic tree on the left was built on the whole conceptual cDNA sequences. On the right side, the same colours under separated regions indicate the same clusters when phylogenetic tree was built for that particular region in figure 4.3.

4.3.6 Dominant BG genes from all four haplotypes belong to BG8-9-12-13 clade with evidence showing selection for variation in cytoplasmic tail but not in other regions

We believed that the dominantly expressed BG gene in T cells from different haplotypes would function the same, and thus could be treated as 'functional alleles'. Therefore, comparing these functional alleles by looking at their sequences at both the nucleotide and amino acid levels, as well as the location and potential clustering of the sequence variations, would give us an insight into the features of these sequences, whether there is evidence of variation for selection, which regions of the BG genes are important for function, etc. In this section, sequence variations were checked for signal sequence, Ig-V domain, transmembrane and cytoplasmic tail by comparing the conceptual transcripts of the dominantly expressed cDNAs (NTBGa, P2aTBGa, 15iTBGa and 6TBGa) from four haplotypes (B21, B19, B15 and B2) to BG8 from B12 haplotype, in order to better understand the questions addressed above.

It is worth explaining the reason for comparing the four genes above to BG8. First of all, as shown already in the phylogenetic trees, NTBGa, P2aTBGa and 15iTBGa are clustered with B12 BG8-9-12 clade, and 6TBGa is identical to BG13, which all together constitute the BG8-9-12-13 clade (Figure 4.2). Second, from the nucleotide sequence alignments between dominants and all 14 BG genes from B12 haplotype (Appendix G), it could be seen that the sequences of NTBGa, P2aTBGa and 15iTBGa are very close to BG8 through whole regions of the cDNA structures, while 6TBGa/BG13 have quite striking differences but are much more closer to BG8-9-12 compared to the rest of the B12 BG genes.

4.3.6.1 5'UTR

Compared to BG8, there are only 15 positions out of 137 nucleotides in the 5'UTR (excluding the primer binding site) that differ in one or another of the dominantly expressed BG genes from the four haplotypes as well as the BG9, 12 and 13 genes of B12 haplotype, and the variation in 5'UTR is of unknown significance (Appendix G).

Figure legend is shown on the page following the figure

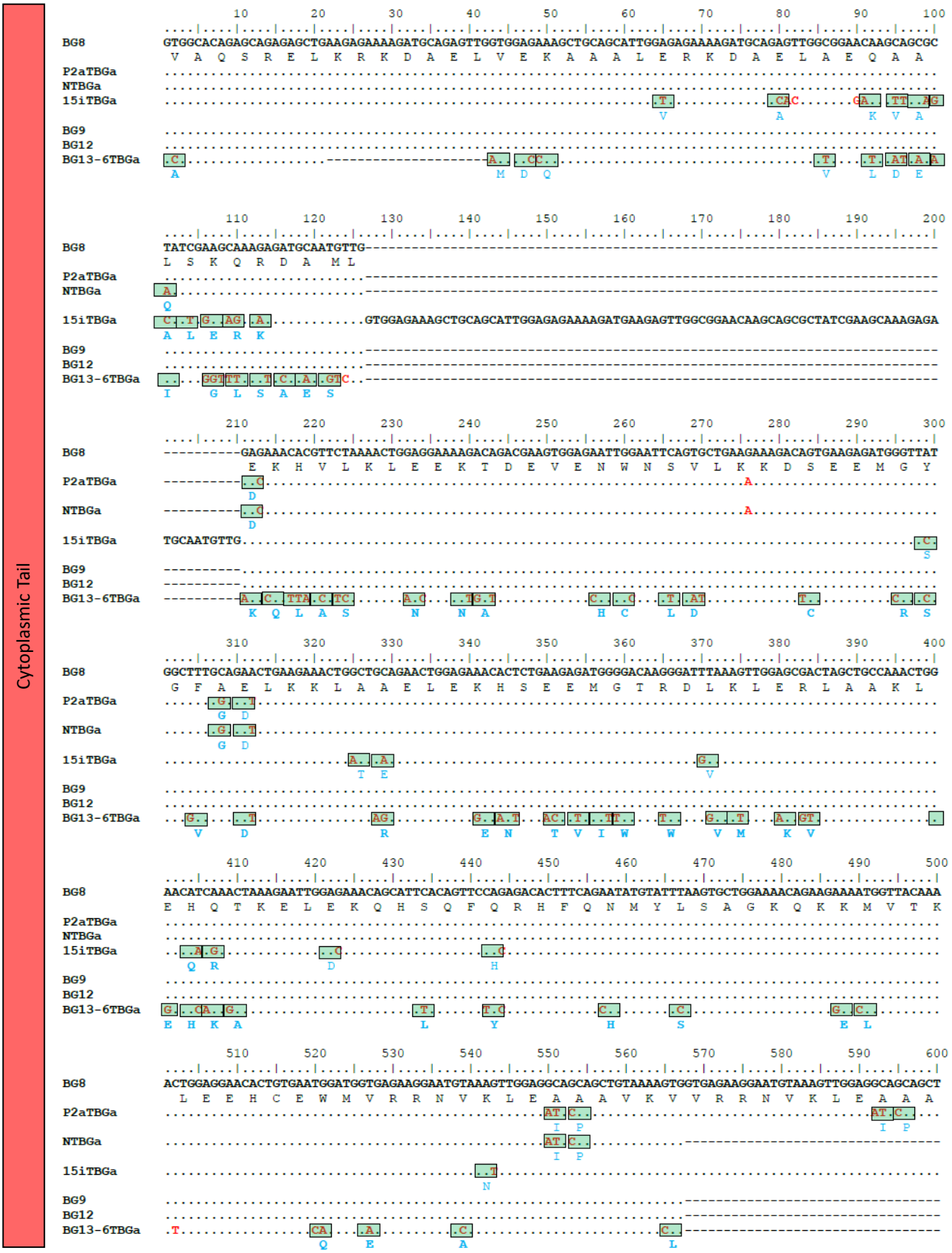


Figure legend is shown on the page following the figure

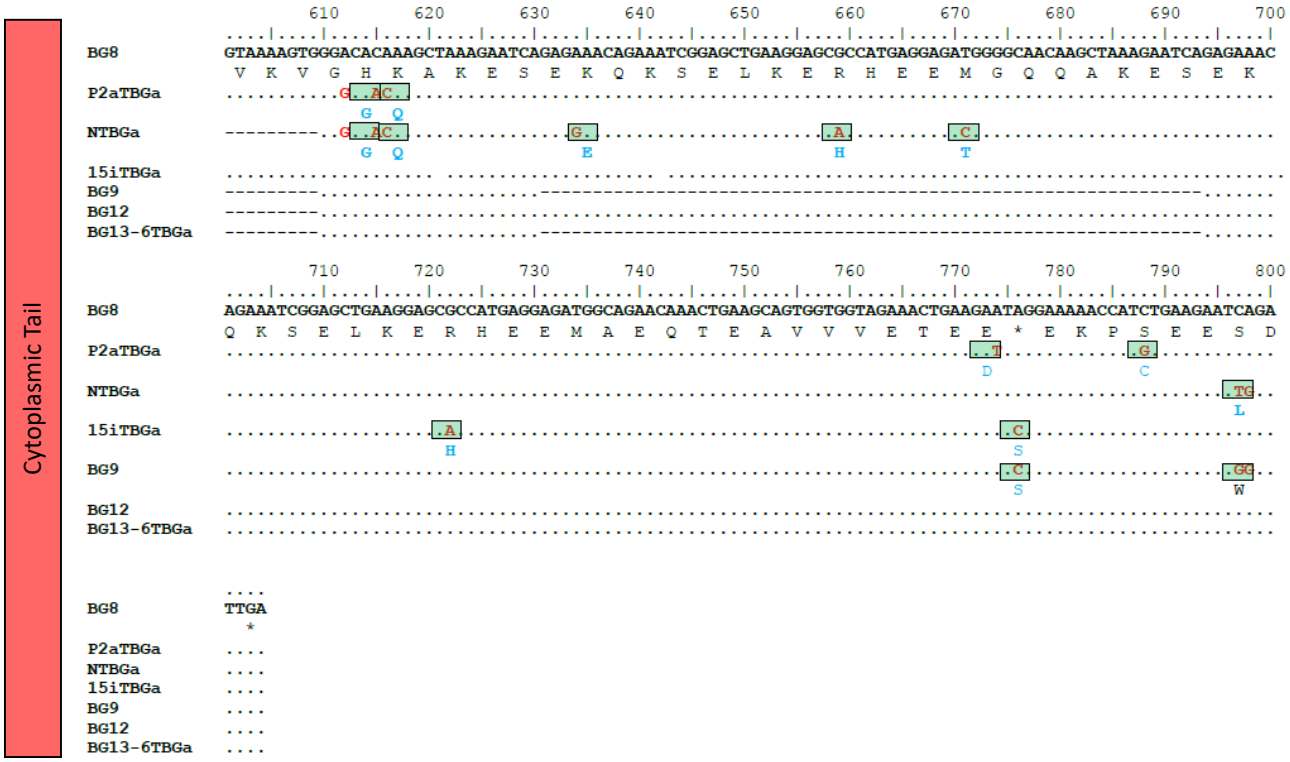


Figure 4.6 Alignments of nucleotide sequences (along with amino acid positions that differ from the BG8 gene from B12 haplotype) for portions of the dominantly-expressed BG genes in T cells from four chicken lines, and for the appropriate BG genes of the B12 haplotype. Names of the transcripts follow the convention: abbreviated line name, "T" for T cells, "BG" and the letter "a" representing the most frequently detected clone from the most frequently detected exon 2 sequence. Portions of the sequence analyzed are: signal sequence from exon 1, exon 2 (mostly Ig-V domain), exon 3 (transmembrane region), and exons corresponding to the cytoplasmic tail (including translated sequences in the last exon). In this figure, the amino acids from split codons at the edges of the exons are assigned to the exon with two of the three nucleotides of the codon (for instance, last amino acid of the signal sequence is assigned to the Ig-V domain, which in fact starts with glutamine in the mature protein); these split codons are not important for tallying up the amino acid differences between the sequences. The four exons apparently inserted into the cytoplasmic tail of P2aBGTa were not considered in this analysis. Codons corresponding to nucleotide changes that lead to amino acid changes are boxed. Dot indicates identity with BG8 sequence; dash indicates not present in all sequences.

		S (1)	S (2)	S (3)	N (1)	N (2)	N (3)	N (1+2)	N (1+3)	N (2+3)	N (1+2+3)
Signal Sequence	NTBGa				2						
	P2aTBGa				1						
	15iTBGa				2						
	BG9				2						
	BG12				2	1					
	BG13-6TBGa			1	2	1		1			
Ig-V Domain	NTBGa			2	1					1	
	P2aTBGa			2				1			
	15iTBGa			3	1	3				1	
	BG9			2	1			1			
	BG12	1		3		1				1	
	BG13-6TBGa			7	3	6			1	2	
TM Region	NTBGa										
	P2aTBGa										
	15iTBGa										
	BG9			1							
	BG12			2				1			
	BG13-6TBGa			3	4	3		1		1	1
Cytoplasmic Tail	NTBGa			2	3	4	3	1		1	
	P2aTBGa			2	3	2	4	2			
	15iTBGa	1		1	4	8	5	2		2	
	BG9						1			1	
	BG12										
	BG13-6TBGa	2			20	16	7	6	4	3	2

Figure 4.7 Compared with BG8 of the B12 haplotype, the number of silent and replacement changes by codon position for the dominantly expressed BG genes in T cells from four chicken lines, and of the other appropriate BG genes from B12 haplotype. Names of the transcripts follow the convention: abbreviated line name, "T" for T cells, "BG" and the letter "a" representing the most frequently detected clone from the most frequently detected exon 2 sequence. Names of the genes follow the convention "BG" and the number of the gene locus from the B12 haplotype. Values are based on the alignments in figure 4.6, and amino acids from split codons at the edges of the exons are assigned to the exon with two of the three nucleotides of the codon (for instance, last amino acid of the signal sequence is assigned to the Ig-V domain, which in fact starts with glutamine in the mature protein).

4.3.6.2 Signal sequence

The whole signal sequences are encoded by exon 1 together with the first two nucleotides from exon 2. Here for the convenience of figure arrangement, the last amino acid of signal sequence, which includes the last nucleotide in exon 1 and the first 2 nucleotides in exon 2, was allocated into Ig-V region in figure 4.6 which doesn't contain any variation, therefore, only the 99 nucleotides were counted for signal sequences.

There are very few differences, 1-6 differences out of 99 nucleotides leading to 0-4 changes in 33 amino acids found in all dominant genes and BG9, 12, 13 compared to BG8. Only one (silent) nucleotide change fails to lead to an amino acid change; therefore, the variation might be selected (Figure 4.7). However, such variation does not change the overall hydrophobic sequence nor does it change the signal sequence cleavage site of three small amino acids.

4.3.6.3 Ig-V domain

Overall, the variation in Ig-V domain compared to BG8 ranges from 4-9 differences out of 342 nucleotides leading to 1-5 changes in 114 amino acids in the three haplotypes (B21, B19 and B15), and 22 nucleotides and 12 amino acids for the B2 haplotype and BG13 (Figure 4.6 and Figure 4.7). The nucleotide changes that lead to no change in the amino acid (silent or synonymous changes) versus those that lead to a change in the amino acid (replacement or nonsynonymous changes) were counted. For three haplotypes, there is no obvious difference in numbers of silent and replacement changes between the dominantly expressed BG gene and BG8 (Figure 4.7). However, there are seven silent and twelve replacement changes compared the dominantly expressed BG gene (6TBGa) in B2 haplotype to BG8. Given that random changes would be expected to lead to twice as many replacement as silent changes, these data are not consistent with strong selection for diversification.

4.3.6.4 Transmembrane region

As discovered already from previous research, there are two types of transmembrane regions, those which have histidine and lysine near the N-terminus of the transmembrane region, and those with a leucine and threonine (Salomonsen *et al.*, 2014). The dominantly expressed BG genes found in this project also fall into the two types. Since 6TBGa is identical to BG13, its transmembrane region bears a lysine; the other three dominant expressed BG genes, together

with BG8, 9 and 12 belong to the other type of transmembrane, which have a threonine instead of the lysine. There is no variation between the transmembrane region sequences of the three haplotypes and only one amino acid difference in BG12. Also, there are only three silent nucleotide changes out of 17 in 6TBGa and BG13, and three codons have multiple nucleotide changes, again consistent with some selection between the BG8-9-12 sequences and the BG13 sequences (Figure 4.6 and Figure 4.7).

4.3.6.5 Cytoplasmic tail

The cytoplasmic tail is composed of amino acid heptad repeats encoded by 21 nucleotide exons (with a few exons of 18 or 24 nucleotides), the numbers of which vary among BG genes. In the alignments between the conceptual cDNA sequences of dominantly expressed BG gene from the four haplotypes and all 14 BG genes from B12 haplotype (Appendix G), there are 33 heptad repeats for BG8. Compared to BG8, BG9 has 28 heptad repeats but a stop codon after a repeat in final exon, giving 29 apparent heptad repeats; BG12 has 31 heptad repeats; BG13 as well as 6TBGa have 27 repeats; NTBGa has 29 repeats; P2aTBGa has 33 repeats; and 15iTBGa has 37 repeats including an apparent insertion of four repeats but in addition a stop codon after a repeat in the final exon, giving 38 repeats. Considering the variation in the numbers of heptad repeats, which would add much complexity into current analysis, only the conserved regions compared to BG8 are taken into account for determination of variation in the nucleotides and amino acids. Much more variation exists in the cytoplasmic tails than in any other region discussed above, and it is striking that most nucleotide changes are nonsynonymous changes (Figure 4.6 and Figure 4.7), indicating the variation is under selection. Detailed examination of cytoplasmic tail features including the amino acid variation is carried out in later section 4.3.8.

4.3.6.6 3'UTR

The 3'UTRs of the dominantly expressed BG genes of all four haplotypes as well as the BG8, 9, 12, and 13 genes from B12 haplotype are co-linear (except for a 20 nucleotide insertion in BG9 that is shared with most BG genes not in the BG8-9-12-13 clade) and nearly identical in sequence (Appendix G). Including the 27 nucleotides that code for protein in BG9 and the dominantly expressed BG gene from B15 haplotype but are untranslated in the other members of this clade, there are only 26 positions out of 411 nucleotides that vary between the eight sequences, with unknown significance.

4.3.6.7 Summary

To summarize, the dominantly expressed BG genes in T cells from three haplotypes (B21, B19 and B15) together with BG8-9-12 from B12 haplotype show evidence of variation for selection in the cytoplasmic tail and maybe in the signal sequence. Between the BG8-9-12 sequences (including the dominantly expressed BG genes from the three haplotypes and BG8-9-12 from B12) and the BG13 sequences (including the dominantly expressed BG gene from line 6_1 and BG13 from B12), strong evidence of variation for selection is seen in the transmembrane region and cytoplasmic tail.

4.3.7 Variation in Ig-V domain structure is mainly located in loop regions

In previous section, variation in the Ig-V domain between dominantly expressed BG genes from the four haplotypes (B21, B19, B15 and B2) and BG8, 9, 12 and 13 from B12 haplotype were compared for silent versus replacement changes, and the conclusion was that the variation is not under selection. However, it is important to examine these changes in detail, including where they are located in the structure.

As shown in figure 4.8, from the amino acid sequence alignments of Ig-V domain between four dominantly expressed BG genes (NTBGa, PTBGa, 15iTBGa and 6TBGa) from four haplotypes (B21, B19, B15 and B2) and BG8, 9, 12, and 13 from B12 haplotype, the amino acid sequences in β-strands are quite conserved with most variation located in other positions. Also there was no change in the cysteines that form the intra-domain disulfide bond, or the cysteine located in the equivalent of complementarity determining region 1 (CDR1) that forms a disulfide bond between the two chains of a BG dimer.

The Ig-V domain structure model of BG8 was built to illustrate the variation. As shown in figure 4.8, the different amino acids from NTBGa, P2aTBGa, 15iTBGa, BG9 and BG12 were labeled in blue, while the differences from 6TBGa and BG13 were labeled in red. From the structure with blue label, it could be seen that nearly all the amino acid variations are located in the membrane distal loops presumably pointing away from the cell surface, with one position in the β-strand and one in the loops underneath the Ig-V domain. In contrast, for 6TBGa and BG13, there is more variation away from the distal loops. Such results might suggest that the dominantly expressed BG genes from T cells of the four haplotypes, together with BG8, 9, 12 and 13 from B12 haplotype, are conserved in the Ig-V structure to maintain

the interaction with their receptor or ligand.

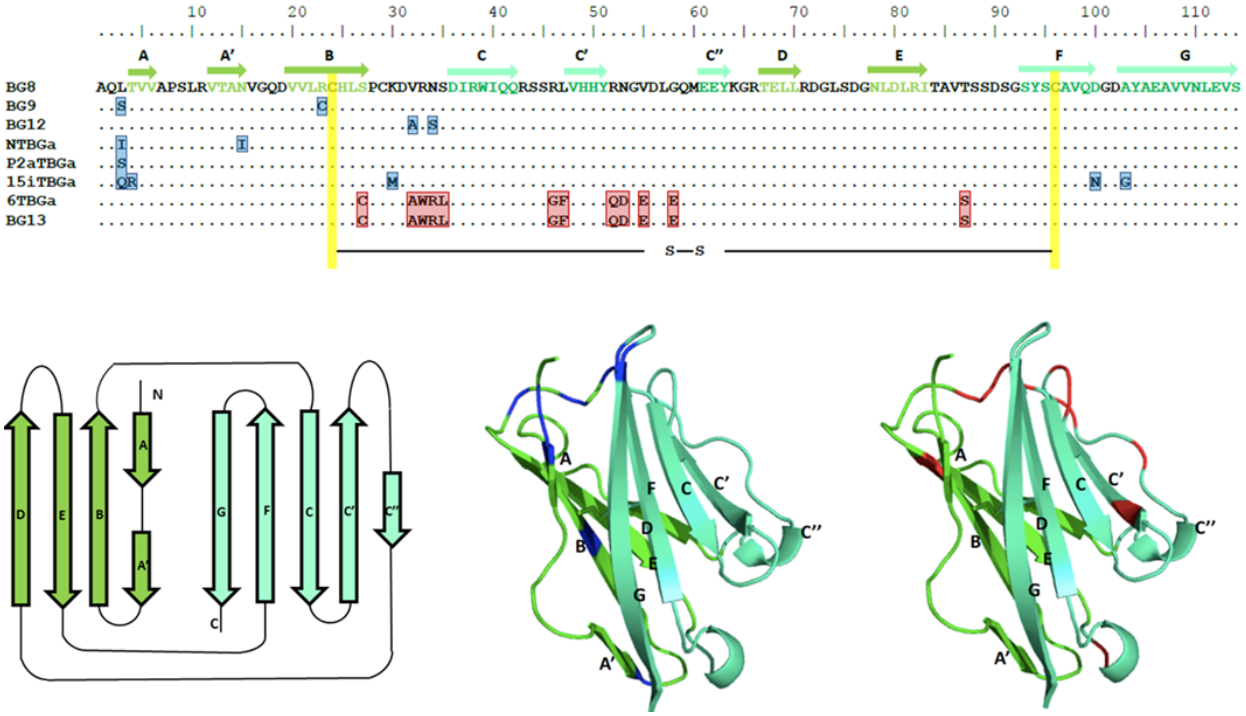


Figure 4.8 Alignment of amino acid sequences for the Ig-V domains of the dominantly expressed genes in T cells from four chicken lines, and for the appropriate genes from B12 haplotype, along with structural models of the Ig-V domains with the location of variation compared with the BG8 sequence of the B12 haplotype indicated. Names of the transcripts follow the convention: abbreviated line name, "T" for T cells, "BG" and the letter "a" representing the most frequently detected exon 2 sequence. Names of the genes follow the convention "BG" and the number of the gene locus from the B12 haplotype. In the top panel, letters indicate amino acids by single letter code, dots indicate identities with BG8 sequence, residues that differ from BG8 are boxed in blue for the three lines, and red for line 6_1 and BG13; yellow indicates the intra-domain cysteines. The β-strands of the V region are indicated by arrows in the top panel, and are colored dark green for one face of the domain and light green for the other face. The same color scheme is used for the three panels below, with the positions of residues for the three lines (and BG9 and BG12) that differ from BG8 colored blue in the middle panel, and positions of residues for line 6_1 and BG13 that differ from BG8 colored red in the right hand panel.

4.3.8 The real organization of heptad repeats in cytoplasmic tail to form coiled-coil was discovered

BG molecules are known to form dimer with two cytoplasmic tails of α-helices forming a coiled-coil (Salomonsen *et al.*, 1987; Kaufman *et al.*, 1990), and in some papers the coiled-coil was also called a leucine zipper (Bikle *et al*., 1996). As described previously, the BG cytoplasmic tails are encoded by many 21 nucleotide exons, and we thought the 1th and 4th codons in the 21 nucleotide exons encode for amino acids that act as the interface between the two chains (Kaufman *et al.*, 1989; Kaufman *et al.*, 1990), with some contribution by the neighbouring amino acids (5th and 7th codons) (Aronsson *et al.*, 2015). However, looking at the exon-intron structure of BG gene carefully, the first amino acid of a heptad is not encoded entirely by this exon but with one nucleotide from the previous exon (Appendix F). Thus the first amino acid encoded by this split codon would vary depending on the previous exon, and this raised the question: is the first amino acid encoded by the 21 nucleotide exon the real first amino acid in the heptad repeat of the α-helical coil?

Therefore, the helical wheels representing cytoplasmic tails of BG8 and BG13 genes were drawn (Figure 4.9), and both two structures clearly showed a pattern that the 4th and 7th codons in one exon encoded amino acids that are mostly hydrophobic, presumably corresponding to the true 1st and 4th amino acids (which from here will be called the *a* and *d* positions) of the true heptad repeats that would form a hydrophobic interface between the two chains. If this hypothesis was true, then in the true *e* and *g* positions we should see some pattern to support the interaction. It could be seen that both in BG8 and BG13 structures, there are many charged amino acids from the first and third codons, potentially allowing salt bridges formed between oppositely charged amino acids of the two chains (Aronsson *et al.*, 2015), which exactly fits the amino acid requirements for *e* and *g* positions. Thus, a clear structure of true heptad repeats forming α-helical coiled coil was determined and it worked well for every BG gene found in this project (Figure 4.10).

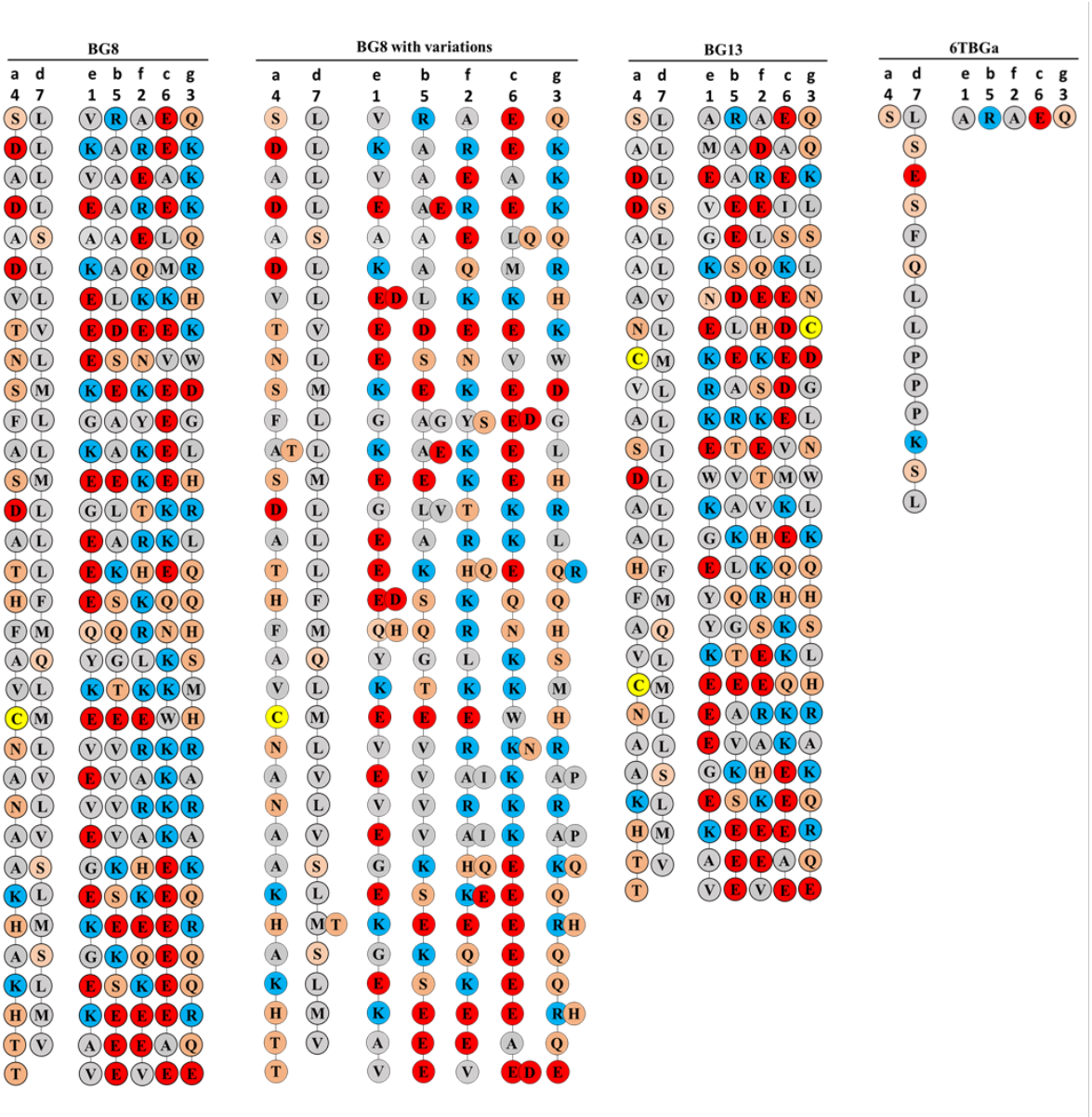


Figure 4.9 Coiled-coil representations of the cytoplasmic tails of BG genes. The first presentation is BG8; the second one is BG8 with the different amino acids found in BG9 and BG12 of the B12 haplotype and in the full-length conceptual transcripts of the dominant sequences of line 15I (B15), line P2a (B19), line N (B21); the third one is BG13 which is identical to the full-length conceptual transcripts of the dominant sequences of line 6_1 (B2); the forth one is the dominant sequence of line 6_1 (B2) with the expected amino acid sequence from the most frequent clone using the real transcript (i.e., exons with the intron read-throughs leading to an early stop codon, sequence 6TBGa-1). The transmembrane region would be at the top of the page, so the C-terminus of the BG protein is at the bottom of the page. The positions of the seven codons in the 21 nucleotide repeat are indicated with numbers at the top, and the position of the seven amino acid positions of the "true heptad repeat" are indicated with letters. Colors of circles surrounding the amino acids (single letter code) indicate features of the amino acids (red, acidic; blue, basic; orange, polar; and gray, hydrophobic except for yellow, cysteine), with the understanding that these features do not correspond to full descriptions of the properties of the amino acids.

At the moment, it is not immediately clear from these structures whether the potential salt bridges favour homodimers or heterodimers with some of the sub-dominantly expressed chains or with the alternatively spliced chains which will be discussed in the coming section. However, the charges in the five positions other than *a* and *d* positions forming the hydrophobic strip between the chains are clustered into acidic, basic and polar patches along the coiled coil, with a particularly clear acidic patch at the C-terminus. Also another striking finding is that the presence of a cysteine residue in the same position of the cytoplasmic tail in the conceptual transcripts of all the dominantly-expressed BG molecules.

It is meaningful to annotate the variation from other dominantly expressed BG genes on the cytoplasmic tail structure of BG8 gene to gain an insight of the potential changes on structure. As shown in figure 4.9, all the variation from three dominantly expressed BG genes (NTBGa, P2aTBGa and 15iTBGa) from three haplotypes (B21, B19 and B15), BG9 and BG12 from B12 haplotype are scattered along the sequence (except for an apparent insertion in the B15 sequence from line 15I), with only one *a* and one *d* position being variable out of 25 variable positions in total. This variation is all di-allelic, most of which is arguably conservative changes (A/T, M/T, E/D, Q/H, A/G, L/V, Y/S, A/I, A/P) with only a few arguably radical changes (A/E, K/E, L/Q, K/N, Q/R, K/Q, R/H). Decorating the coiled coil representation of the cytoplasmic tail sequence revealed that much of the variation is located in two parts of the coil, 11-18 and 23-26 of 33 heptads, but whether this constitutes clustering is not yet clear. The cytoplasmic tails from the dominantly expressed conceptual transcript of line 6_1 (B2) and from the BG13 gene (B12) are shorter than all the genes described above. And interestingly, the real cytoplasmic tail of the dominantly expressed gene in line 6_1, 6TBGa, is much shorter due to alternative splicing, which will be discussed in the coming section.

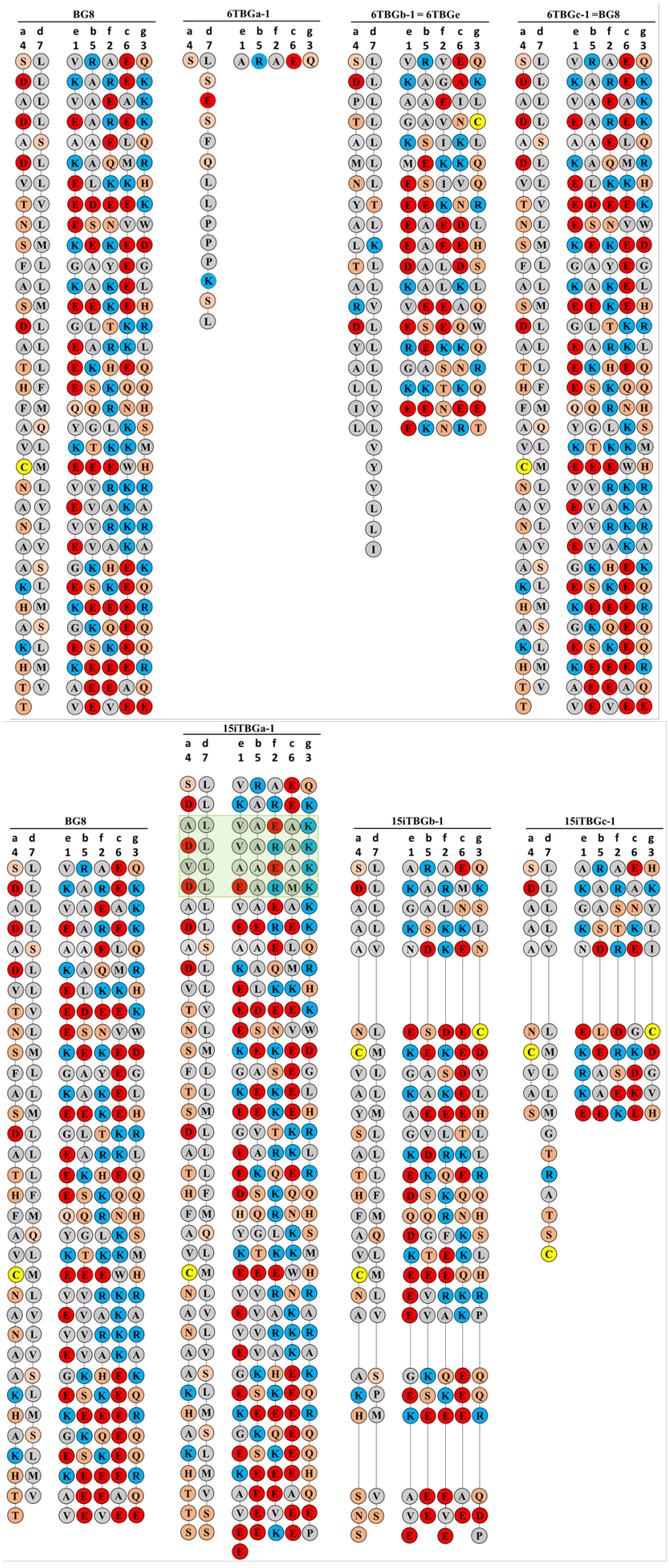


Figure legend is shown on the page following the figure

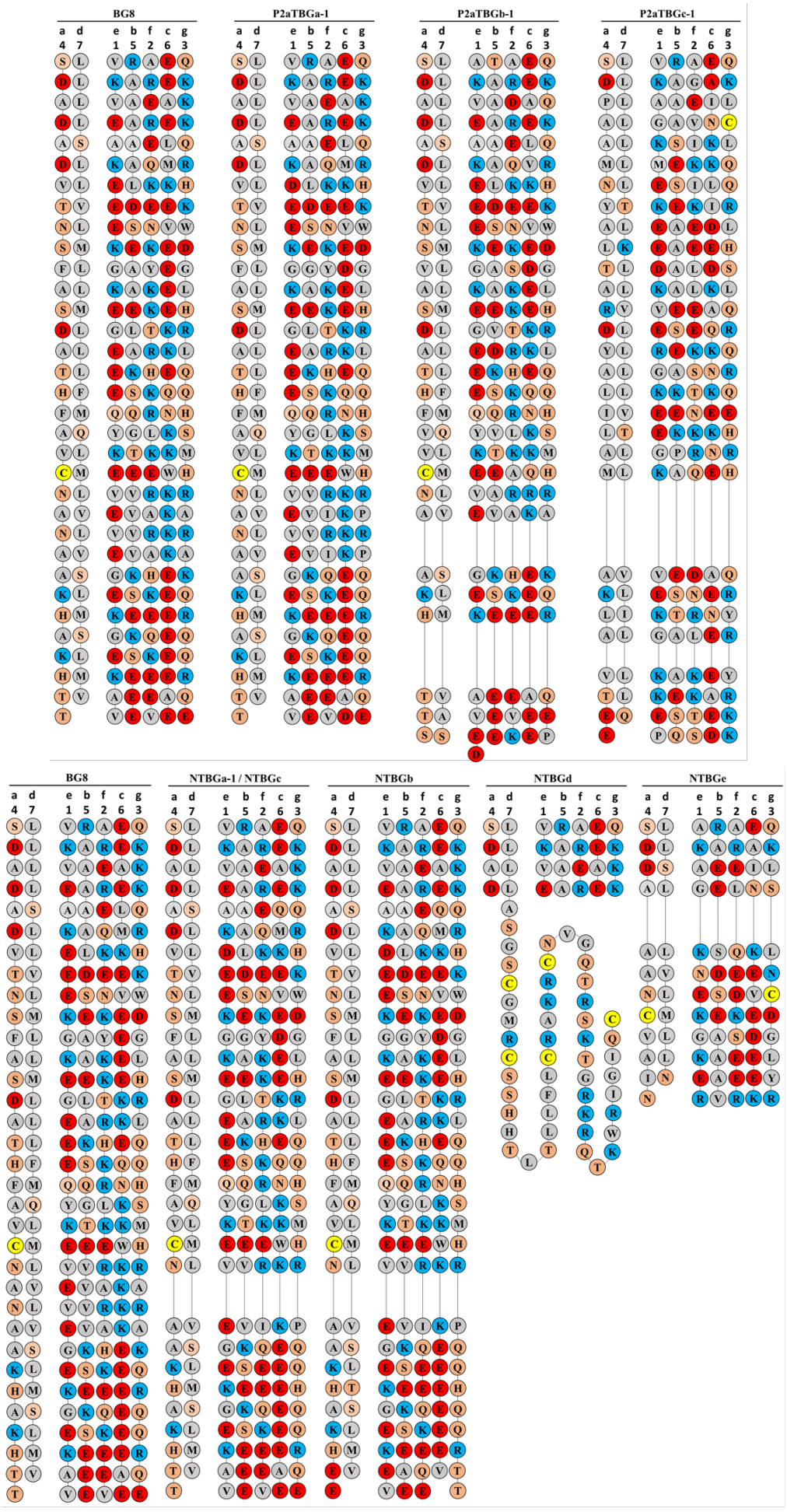


Figure legend is shown on the page following the figure

Figure 4.10 Coiled-coil representations of the cytoplasmic tails inferred from the most frequent clones for each of the 16 genes found in T cells from four chicken lines using the real transcripts. Names of the transcripts follow the convention: abbreviated line name, "T" for T cells, "BG" and the letter "a" representing the most frequently detected clone from the most frequently detected exon 2 sequence (and "b" representing the most frequently detected clone from the second most frequently detected exon 2 sequence, and so forth), followed in some cases by a dash and then a number representing the alternative splicing variant with "1" being the most frequently detected clone. The transmembrane region would be at the top of the page, so the C-terminus of the BG protein is at the bottom of the page. The positions of the seven codons in the 21 nucleotide repeat are indicated with numbers at the top, and the position of the seven amino acid positions of the "true heptad repeat" are indicated with letters. Colors of circles surrounding the amino acids (single letter code) indicate features of the amino acids (red, acidic; blue, basic; orange, polar; grey, hydrophobic except for yellow, cysteine), with the understanding that these features do not correspond to full descriptions of the properties of the amino acids. The amino acids encoded by the four extra exons inserted in 15iBGTa are indicated by a light green box, and elsewhere the positions of the heptads are shifted to illustrate sequence similarities or identities between cytoplasmic tails.

4.3.9 Lots of alternative splicing in the cytoplasmic tails were observed in most BG genes

So far, the analyses above are mainly based on conceptual cDNA transcripts, meaning no introns retained in the sequence. However, the most striking finding in this project is that in real BG cDNA transcripts, lots of alternatively spliced isoforms were observed for most of the BG genes. Nucleotide alignments of 57 cDNA sequences (Appendix D) have shown a variety of different alternative splicing in cytoplasmic tails. Figure 4.11 summarized the intron-exon structure of 27 cDNA transcripts against the genomic sequence of BG8 from B12 haplotype; these cDNA transcripts were found in at least two PCRs (except some subdominant expressed genes from line N) and their sequences were submitted to GenBank with the accession numbers listed in appendix H. All of the retained introns lead to in-frame stop codons, some of which are located long before the stop codon expected from the conceptual transcripts (Figure 4.11). The cDNA transcripts of the dominantly expressed BG gene from line 6_1, 6TBGa, retain the intron directly after the first 21 nucleotide exon, which truncates the cytoplasmic tail after only 13 amino acids. This truncated cytoplasmic tail differs greatly from the dominantly expressed BG genes from the other three haplotypes which all have long cytoplasmic tails in their dominantly expressed cDNA transcripts (Figure 4.11). Some of the subdominant sequences also have truncated cytoplasmic tails, some with clusters of cysteines (Figure 4.11), for which the potential functions are not clear at the moment.

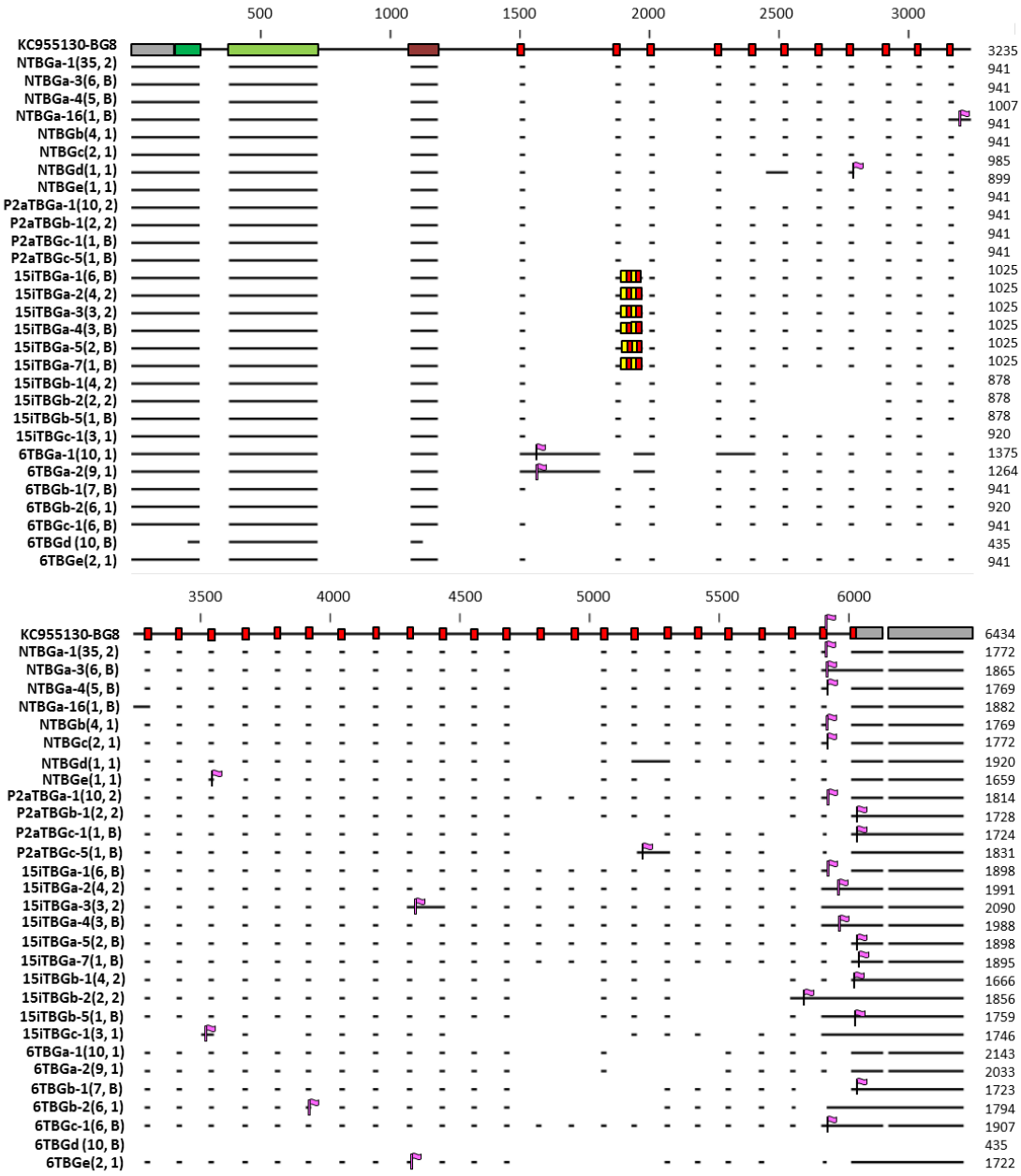


Figure 4.11 Representation of the intron–exon structure of the BG genes inferred from 29 cDNA sequences of the 16 BG genes found in T cells of four chicken lines. The 29 sequences include the dominant transcript found in each of the 16 BG genes, as well as those alternatively expressed transcripts identified in more than one independent PCR, indicating the actual mRNA transcripts as horizontal lines and stop codons as vertical purple flags. In 15iTBGa, the alternating red and yellow boxes indicate the four additional heptad repeats found for this gene, compared with BG8 (B12). Names of the transcripts follow the convention: abbreviated line name, "T" for T cells, "BG" and the letter "a" representing the most frequently detected clone from the most frequently detected exon 2 sequence (and "b" representing the most frequently detected clone from the second most frequently detected exon 2 sequence, and so forth), a dash and then a number representing the alternative splicing variant with "1" being the most frequently detected clone (and "2" the second most frequently detected clone, and so forth). Numbers in parentheses indicate the number of clones found for a particular full sequence, followed by the number of independent PCRs in which the sequence was identified (1, found in one PCR; 2, found in two PCRs; B, found in one PCR described in this chapter and one using B cell cDNA, data shown in chapter 5). The 29 sequences were deposited in GenBank, with the accession numbers listed in appendix H.

4.3.10 Summary

As summarized in figure 4.12, in total 16 different BG genes were found from peripheral T cells of four different chicken lines, line N (B21), line P2a (B19), line 15I (B15) and line 6_1 (B2). Only one dominantly expressed BG gene was found in each line, together with significant levels of sub-dominant genes. Most of these genes were clustered to BG8-9-12 clade when compared to all 14 BG genes from B12 haplotype, most with type 2 cytoplasmic tail and 3'UTR, and some with type 1 cytoplasmic tail and 3'UTR. Also, alternatively spliced isoforms mainly due to intron retention leading to truncated cytoplasmic tails were found for most of these transcripts, particular for the dominantly expressed gene in line 6_1, 6TBGa. 6TBGa has identical conceptual sequence to BG13 gene from B12, but the cDNA transcripts found in this project contain an intron just after the first exon in cytoplasmic tail, leading to a very short tail. The dominantly expressed BG genes from the other three lines were NTBGa, P2aTBGa and 15iTBGa, which are all closely related to BG8, 9 and 12 genes from B12 haplotype, with some nucleotide variation in sequence throughout the whole cDNA regions, and with evidence for selection in cytoplasmic tail but not for the Ig-V domain.

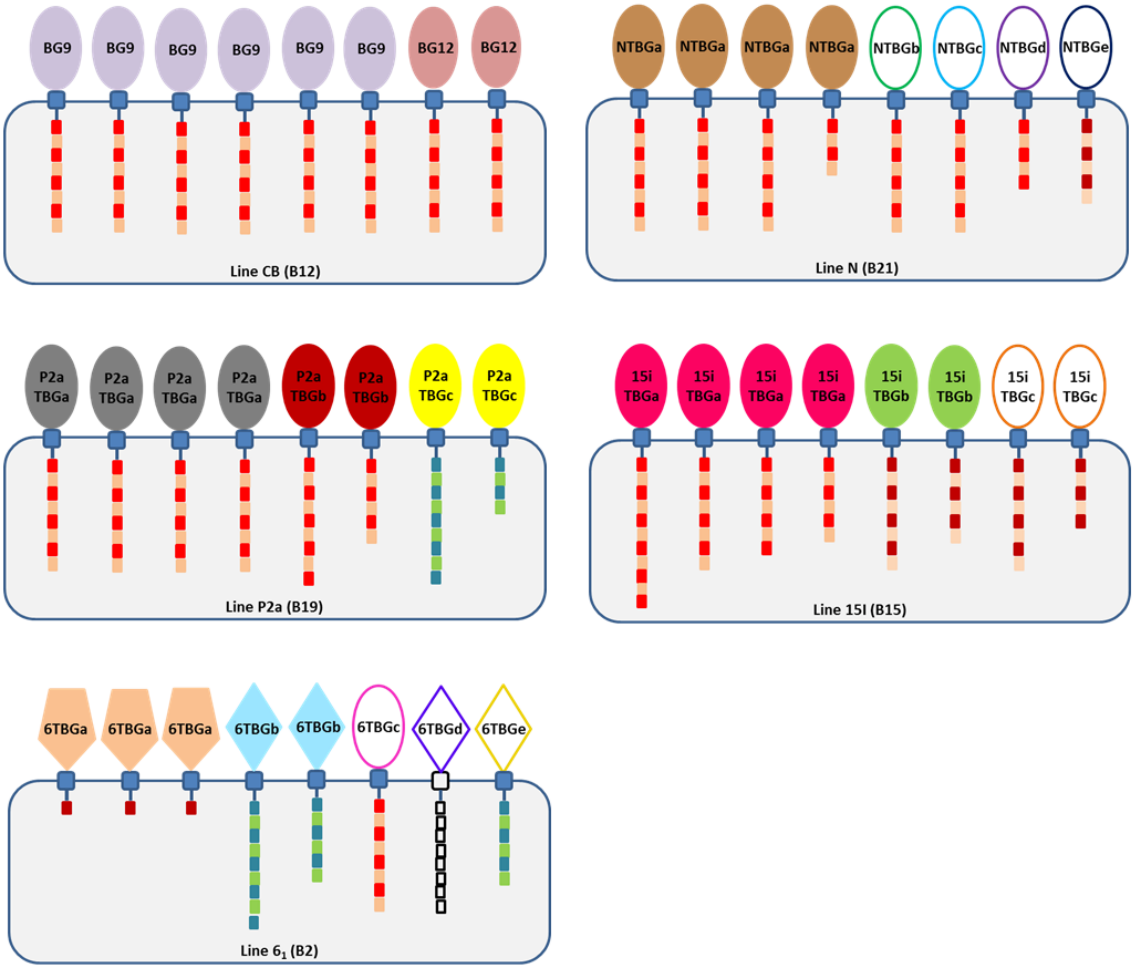


Figure 4.12 Cartoon summary of the findings about BG transcripts in T cells from four chicken lines. There are no expressed sequences identical between the four chicken lines but line 6_1 (B2) has sequences identical to genes of the B12 haplotype (which, however, are not expressed in T cells of the B12 haplotype). The most expressed sequences are from the BG8-9-12 clade although line 6_1 has a dominantly expressed BG sequence of the BG13 clade and two subdominant sequences from the BG5-7-11 clade. The cytoplasmic tails are mostly type 2a and the length varies due to alternative splicing (intron read-through). The cartoon shows BG proteins in cells of each haplotype, with the numbers of each protein reflecting the ratio of different sequences in that haplotype. Extracellular Ig-V domains are represented by shapes to indicate relationship to clades of BG genes from the B12 haplotype (ovals, BG8-9-12 clade; pentagons, BG13 clade; and diamonds, BG5-7-11 clade) and by colour (colours as in Figure 2, with those sequences found in only one PCR represented by ovals and diamonds not filled with colour); cytoplasmic tails indicated by boxes representing heptad repeats, with lengths correlated with the length of the tail taking into account alternative splicing and with colours representing the clade (type 1, blue and green; type 2a, bright red and brown; type 2b, dark red and brown; and 6TBGd, white since no data is available).

4.4 Discussion

4.4.1 What did we learn in this project?

This is the first time that BG genes from purified T cells were examined systematically from four different chicken lines, which were line 6_1, line 15I, line N and line P2a, with the MHC haplotypes B2, B15, B19 and B21, respectively. The raw data were very complicated; even though we already knew BG genes were polymorphic, we didn't expect so many different sequences. The idea of using exon 2 (encoding mainly the Ig-V domain without signal sequence) to distinguish different BG genes successfully identified 16 BG genes expressed in T cells from four chicken lines above, which made us realize that many complicated cDNA sequences were due to alternative splicing in the cytoplasmic tail.

Examining these alternatively spliced transcripts carefully, we found that they are mainly due to intron retention (also called intron read-through), which shortens the cytoplasmic tail compared to what is expected from the conceptual gene sequence, for instance, dramatically truncating nearly the whole cytoplasmic tail of the dominantly-expressed BG gene of line 6_1 (B2). Such intron read-through, a form of alternative splicing, was first noticed long ago in the cytoplasmic tails encoded by BG cDNAs (Kaufman *et al*., 1989, 1990). Some intron read-through seems to have become fixed, for example BG1 genes in which an active immunotyrosine inhibition motif (ITIM) is located in an exon bounded by two 21 nucleotide repeats (Goto *et al*., 2009; Chattaway *et al*., 2016). A bioinformatic analysis of the 14 BG genes of the B12 haplotype found many read-through introns that led to in-frame stop codons, but no additional signaling motifs were obvious in those introns that read through in-frame (Salomonsen *et al*., 2014). In our project, a large number of alternatively spliced transcripts are observed which deserves further study. For example, we could transfect cell lines with these alternative splicing isoform clones and see whether they can be expressed at least in vitro.

By drawing helical wheels of the cytoplasmic tails, we realized that the true *a* and *d* positions in a heptad repeat forming the α-helical coiled coil are not encoded by the first and forth codons in the 21 nucleotide exon, but the fourth and seventh codons. This is sensible, since if we look at the intron-exon structure of well characterized BG genes from B12 haplotype carefully, the first amino acid of a heptad is not encoded entirely by this exon but with one nucleotide from the previous exon. Therefore, the first amino acid in a heptad repeat is

encoded by split codon, and it would vary depending on the previous exon. Having realized the likely structure of coiled-coils in BG proteins, we noticed many patterns organized by certain amino acids in some regions. However, further research needs to be carried in order to understand such findings.

The most important information we were expecting from this project was to identify 'functional alleles' of BG genes from T cells of different haplotypes. By comparing the sequences of these functional alleles, we wanted to understand whether variation in the extracellular Ig-V domain and/or the cytoplasmic tail showed evidence for selected function. We found that the dominantly-expressed BG genes from all four haplotypes come from one clade of BG genes in B12 chickens (BG8-9-12-13, haemopoietic 5' UTR with type 2 cytoplasmic tail and 3'UTR). Among them, three dominantly expressed BG genes from three haplotype (B21, B19 and B15) are closed related to BG8-9-12, while the other dominantly expressed BG gene from the B2 haplotype, 6TBGa, is identical to BG13 in the conceptual sequence. However, the mystery is that in the B12 haplotype, BG13 was not detected in T cells at all by using SS-TM primers.

Comparing the nucleotide and amino acid variation in these dominantly expressed BG genes from the four haplotypes with BG8, 9, 12 and 13 genes from B12 haplotype using their conceptual cDNA sequences, led to the following findings.

For the Ig-V domain, low levels of variation were found. The amino acids in the β-strands are quite conserved within the dominantly expressed BG genes from the four haplotypes and BG8, 9, 12 and 13 from B12 haplotype. The variation was mainly found in the distal loops except the variation from BG13, suggesting selection for functional interactions with other molecules. However, there is no evidence of selection for variation based on silent versus replacement changes; perhaps data from additional haplotypes will help.

For the cytoplasmic tail, there was clear support for selection of variation. The variation in the cytoplasmic tail for the conceptual transcripts from the three haplotypes with similar BG sequences (and the BG8, BG9 and BG12 genes from the B12 haplotype) is predominantly located in two stretches at the five positions that are not in the hydrophobic stripe between the two α-helices of the coiled coil. The functional significance of this variation is not clear, but the evidence from silent versus replacement substitutions supports selection for this variation. The only two known examples of function for BG molecules, regulation of actin-myosin by

"zipper protein" in intestinal epithelial cells and effect of the BG1 gene on viral disease (Bikle *et al.*, 1993; Bikle *et al.*, 1996; Goto *et al.*, 2009), are both associated with the cytoplasmic tail rather than the extracellular Ig-V domain, which may fit with the notion that the cytoplasmic tail is under selection for variation.

4.4.2 The limitations

Two potential limitations of the work represented in this chapter might be that PCR has a certain level of artifact, and whether the colony numbers really stand for the expression level of a gene. First of all, we have done two independent PCR amplifications to rule out the bias; for line 6_1 an extra PCR was carried out using another primer pair. Second, we had tried our best to pick as many as 92-96 clones for colony PCR to select positive clones for sequencing which was overweigh to our previous research (Chattaway, 2013). However, for these transcripts which only appeared in one PCR reaction, we remain suspicions and did not use them for critical analysis.

Another argument might be that we cannot rule out the possibility that the transcripts with intron read-through might be incompletely spliced nuclear RNAs which would never be translocated to the cytoplasm or be translated. However, the RNA was primed with olig-dT for the reverse transcription step and the amplicons are nearly full-length, so it seems most likely that these RNAs are polyadenylated.

We have tried our best to avoid artifacts occurring in our data, however, such concerns above should not be ignored. In the future, more sophisticated tools, for example single-cell RNA-seq, could be carried out to evaluate our results. Also, equivalent research on protein level is required to ensure that these transcripts with alternative splicing in cytoplasmic tails encode real BG proteins.

Chapter 5

Functional alleles of chicken BG genes in peripheral B cells

5.1 Introduction

Previously in the T cell project, sixteen BG genes were found from T cells of four different chicken lines, line N (B21), line P2a (B19), line 15I (B15) and line 6_1 (B2) (Chen *et al.*, 2018). The improvement of methods, including using HU primers, distinguishing BG genes by exon 2 sequence etc., has allowed us to better understand some important features of BG genes at molecular level. For example, lots of alternative splicing isoforms of BG cDNA sequences for most BG genes were found, which helped to explain why BG proteins were seen with different sizes (Kaufman *et al.*, 1990). The helical wheel structures of the cytoplasmic tails revealed the true positions of each amino acid from the heptad repeats located in the coiled-coil. The main purpose for the T cell project was to find 'functional alleles'. The dominantly expressed BG genes found in T cells from different chicken lines were expected to have the same function, and thus they could be treated as functional alleles. Through comparing these dominantly expressed BG genes, the BG8-9-12 alleles were found for three lines, but not for the dominantly expressed BG gene from line 6_1, 6TBGa, which was found to be identical to BG13 in B12 haplotype. Further analysis of these BG8-9-12 alleles showed evidence of amino acid variation for selection in the cytoplasmic tail. Such a successful study leads us to repeat the same work on B cells, trying to understand the following questions.

How many BG genes can we find from B cells in these four chicken lines? What do the phylogenetic relationships of these BG genes look like? Is alternative splicing also seen in the cDNA of most BG genes found in B cells, and if so, what types of alternative splicing are these transcripts?

Which BG genes are the functional alleles in B cells? By comparing these allelic BG genes, which regions of their cDNA sequences are conserved, and is there evidence of variation for selection in these regions?

Comparing the BG genes found in B cells to the ones that were found previously in T cells, what are the differences and similarities? Comparing all the BG genes found in both T and B cells from our project to the other BG cDNA sequences from GenBank or ENSEMBL, what are the relationships of these genes? Do chicken lines with the same B haplotype have the same BG genes? Do chickens from a particular line all have the same sequence of BG genes? What are the relationships of BG genes among many different chicken lines?

5.2 Materials and methods

5.2.1 Chicken lines, haplotypes and samples

The chicken lines, haplotypes and samples are the same as those in T cell project in section 4.2.1 chapter 4 except the antibody used for B cell sorting was Bu-1a-RPE (Southern Biotech).

5.2.2 RNA isolation, cDNA synthesis and PCR amplification

The protocols for RNA isolation, cDNA synthesis and PCR are the same as those in T cell project in section 4.2.2 (chapter 4).

5.2.3 Cloning and sequencing

The cloning and sequencing procedures are the same as those in T cell project in section 4.2.3 chapter 4 except the sequencing primers, which were T7, pJETR, UC699, UC700, UC701, UC702 and UC703, with the detailed oligo sequences listed in appendix B.

5.2.4 Sequence analysis

Sequence analysis method is the same as that in T cell project in section 4.2.4 chapter 4.

5.2.5 BG sequences from other sources

5.2.5.1 Fourteen BG genes from B12 haplotype

All the 14 BG genes of B12 haplotype were well annotated in GenBank with the accession number of KC955130.1 (Salomonsen *et al.*, 2014).

5.2.5.2 BG genes annotated from chicken whole genome sequence

In order to compare BG gene sequences from as many different haplotypes as possible, the latest version of chicken whole genome sequence (WGS), Gallus_gallus-5.0 (Gaga5.0 for short) with GenBank Assembly Accession number of GCA 000002315.3, was annotated within ENSEMBL by BLAT using the exon 2 nucleotide sequence of the BG8 gene from the B12 haplotype. Gaga5.0 was assembled from sequencing libraries prepared from DNA of a single female of an inbred line of red jungle fowl (UCD001) (Hillier *et al.*, 2004). The MHC

haplotype of line UCD001 was BQ, which is closely related to the standard B21 haplotype found in experimental lines of chickens derived from egg layers (Senseney *et al.*, 2000).

There were two reasons to use the exon 2 sequence to search for BG genes. Firstly, exon 2 encoding the Ig-V domain is the only region that is long enough in sequence (342 bp) and consistent in size and present in all BG genes. Exon 1 encodes 5' UTRs and signal sequences, but the 5' UTRs vary between haemopoietic BG genes and tissue BG genes. Exon 3 encoding transmembrane region is too short (105 bp) for specific searching; the cytoplasmic tail region is composed of many 21 nucleotide exons (with a few exons of 18 or 24 nucleotides), and varies among different BG genes. The 3' UTRs also vary between type 1 and type 2 BG genes. Secondly, since exon 2 has worked very well for defining BG genes in previous T cell project for four different chicken haplotypes, it should also work for the BQ haplotype.

5.2.5.3 cDNA sequences downloaded from GenBank

A total of 24 BG cDNA sequences was downloaded from GenBank database into a local CLC file folder. 'BG chicken' was used as the key word to search GenBank database; all the cDNA sequences showing BG genes were checked in detail and only those coding DNA sequence (CDS) which included the full exon 2 region were downloaded. Detailed information of the 24 BG cDNA sequences is summarized in table 5.1. Arabic numbers listed in the first column are used to indicate each sequence for convenience. The gene names in the second column consist of the abbreviations for each sequence's GenBank accession number and the haplotype/chicken information specified by the authors submitting the sequences. Among the 24 BG cDNA sequences, seven (number 1 to 7) are full-length cDNA sequences while the rest were partial regions of BG genes, with one (number 8) containing an incomplete Ig-V domain, full transmembrane region and incomplete cytoplasmic tail, and the other 16 sequences having the Ig-V domains with incomplete signal sequences.

It is also important to understand from which chicken lines and MHC haplotypes these sequences were derived. The sequences from number 1 to 6 with B21 haplotype were sequenced from cDNA library made using erythroid cells of UCD330 line chicken embryos. According to the author, UCD330 line was highly inbred and homozygous for the B21 haplotype (Miller *et al.*, 1991). The sequence number 7 (U49098.1) was unspecified for the haplotype or chicken line. The author only mentioned that it was sequenced from the RACE (Rapid Amplification of cDNA End) products using the mRNA template directly generated

from the intestinal mucosa of white leghorn cockerels (Bikle *et al*., 1996). The sequence number 8 (U32560.1) was also unspecified for the haplotype or chicken line. In 1985, Moon et al. made a λgt11 expression library; however, the author didn't specify from which chicken line or haplotype it was generated but only with the description of 'erythroid cells from 14-15-d-old chicken embryos' (Moon *et al.*, 1985). Later, Goto et al. obtained a phage DNA clone called λbg28 from this library, found this clone was BG positive with antiserum prepared against purified B-G21 antigen (B-G21 referred to BG alleles from chicken line UCD330 which is MHC B21 haplotype) (Goto *et al.*, 1988), and sequenced this clone, which is the sequence number 8. The rest of the sequences with Camperos noted in the column of 'line' in table 5.1 were all submitted by Miller's group. For the sequences from number 9 to 20, there was no detailed information about chicken haplotype and the only information we could get was that the sequences were from Camperos chickens. For the sequences from number 21 to 24, a paper published by the authors specified that they were from 51 randomly picked Camperos which were free-range broiler chickens at the National Institute of Agricultural Technology (Pergamino, Argentina) (Iglesias *et al.*, 2007).

Table 5.1 The summary of all 26 BG cDNA downloaded from GenBank

	Gene name	GenBank ID	Haplotype	Line	Length	Definition	PubMed	date
1	M2-B21	M61862.1	B21	UCD330	1429 bp	Chicken B-G mRNA, 3' end.	1903541	1991
2	M4-B21	M61864.1	B21	UCD330	1808 bp	Chicken B-G mRNA, complete cds.	1903541	1991
3	M3-B21	M61863.1	B21	UCD330	1826 bp	Chicken B-G mRNA, complete cds.	1903541	1991
4	M1-B21	M61861.1	B21	UCD330	1741 bp	Chicken B-G mRNA, complete cds.	1903541	1991
5	M0-B21	M61860.1	B21	UCD330	1790 bp	Chicken B-G mRNA, complete cds.	1903541	1991
6	NM-B21	NM_001030673.1	B21	UCD330	1119 bp	Gallus gallus V-region-like B-G antigen (V-BG), mRNA.	1903541	1991
7	U98-Unk	U49098.1			1777 bp	Gallus gallus zipper protein mRNA, complete cds	8621557	1996
8	U60-Unk	U32560.1			523 bp	Gallus gallus MHC B complex B-G subregion mRNA, partial cds.	2826332	1996
9	DQ49-Cam	DQ176449.1		Camperos	399 bp	Gallus gallus clone G8 MHC class IV antigen (B-G) gene, exons 1, 2	-	2016
10	DQ48-Cam	DQ176448.1		Camperos	399 bp	Gallus gallus clone G7 MHC class IV antigen (B-G) gene, exons 1, 2	-	2016
11	DQ47-Cam	DQ176447.1		Camperos	399 bp	Gallus gallus clone G6 MHC class IV antigen (B-G) gene, exons 1, 2	-	2016
12	DQ46-Cam	DQ176446.1		Camperos	399 bp	Gallus gallus clone G5 MHC class IV antigen (B-G) gene, exons 1, 2	-	2016
13	DQ45-Cam	DQ176445.1		Camperos	399 bp	Gallus gallus clone G4 MHC class IV antigen (B-G) gene, exons 1, 2	-	2016
14	DQ44-Cam	DQ176444.1		Camperos	399 bp	Gallus gallus clone G9 MHC class IV antigen (B-G) gene, exons 1, 2	-	2016
15	DQ43-Cam	DQ176443.1		Camperos	399 bp	Gallus gallus clone G3MHC class IV antigen (B-G) gene, exons 1, 2	-	2016
16	DQ81-Cam	DQ903881.1		Camperos	399 bp	Gallus gallus clone 3G MHC class IV antigen (B-G) gene, exons 1, 2	-	2016
17	DQ80-Cam	DQ903880.1		Camperos	399 bp	Gallus gallus clone 1G MHC class IV antigen (B-G) gene, exons 1, 2	-	2016
18	DQ82-Cam	DQ903882.1		Camperos	399 bp	Gallus gallus clone 6G MHC class IV antigen (B-G) pseudogene,	-	2016
19	DQ42-Cam	DQ174444.1		Camperos	399 bp	Gallus gallus clone G2 MHC class IV antigen (B-G) gene, exons 1, 2	-	2016
20	DQ41-Cam	DQ174443.1		Camperos	399 bp	Gallus gallus clone G1f MHC class IV antigen (B-G) gene, exons 1, 2	-	2016
21	AF72-Cam	AF388372.1		Camperos	399 bp	Gallus gallus clone G.4.1.e MHC type antigen B-G (B-G) gene, exons	12648091	2016
22	AF71-Cam	AF388371.1		Camperos	399 bp	Gallus gallus clone G.4.2.5 MHC type antigen B-G (B-G) gene, exons	12648091	2016
23	AF70-Cam	AF388370.1		Camperos	399 bp	Gallus gallus clone G.4.1.2 MHC type antigen B-G (B-G) gene, exons	12648091	2016
24	AF69-Cam	AF388369.1		Camperos	399 bp	Gallus gallus clone G.4.1.c MHC type antigen B-G (B-G) gene, exons	12648091	2016

5.3 Results

5.3.1 HU PCR results and the nomenclature of BG sequences

The same methodology and procedure used for detecting BG genes in T cells was applied to the purified peripheral B cells from the same individuals of four chicken lines, line N (B21), P2a (B19), 15I (B15) and 6_1 (B2). Therefore, the detailed information regarding how BG genes were found, how to handle the sequencing data, how to organize the sequences etc., is omitted in this chapter.

The nomenclature of BG sequences followed the same principle for T cell project in general, but to avoid any confusion, two important points should be clarified. First of all, the BG genes (based on exon 2 sequences) found previously in T cell project and also found in B cell project kept the gene name with only replacement of 'T' to 'B' in order to specify such genes were found from B cell project. For example, NBBGa gene is the same gene as NTBGa gene. Also NBBGa, NTBGa, NTBBGa and NBGa all refer to the same gene, with the exact same conceptual cDNA sequence (that is, exons without any intron retention). The reason to use different nomenclature to refer to the same gene is for the convenience of understanding the cellular source of the gene in different contexts. However, when it comes to the clone name or sequence name, situation was different. For example, the cDNA sequence of NBBGa-1 might or might not have the same nucleotide sequence as NTBGa-1(usually due to differences in splicing of the cytoplasmic tail exons), although they were both the most dominantly expressed cDNA of NBGa gene in B cells and T cells, respectively. Second, if such genes found in B cells were not found in previous T cell project, the gene was named following the sequence of the letter after the last letter used in T cell project. Taking line N as an example again, in T cell project there were five BG genes found, named from NTBGa to NTBGe, while in B cell project, an extra BG gene was found; therefore it was named 'NBBGf', just after the last letter 'e' used in T cell project.

5.3.2 Eighteen BG genes were found from B cells of four chicken lines

In total, there were 18 different BG genes (based on exon 2 sequences) found from sorted peripheral B cells of four chicken lines, line N (B21), line P2a (B19), line 15I (B15) and line 6_1 (B2). BG gene numbers varied in different lines, ranging from two in line N to eight in line 6_1.

5.3.2.1 Two BG genes were found in line N (B21)

In line N (B21), only two BG genes, NBBGa and NBBGf, were found in B cells with NBBGa dominantly expressed. During the first two independent PCRs using HU primers, NBBGa was dominantly expressed, while NBBGf was found only in the second PCR. To further evaluate this result, a third PCR using SS-TM primers confirmed that there were only these two genes expressed with NBBGa being dominant (Table 5.2 and Figure 5.1). As shown in table 5.2, there are as many as 22 different cDNA transcripts found for NBBGa with one transcript found by SS-TM primers only showing partial sequence (partial signal sequence, the whole Ig-V domain and partial transmembrane region). Some of these transcripts appeared in both PCRs, but some in one PCR with some even having only one clone. However, it was hard to rule out many of these sequences with only one clone (except the ones with obvious PCR errors), because some of them were proved to exist in previous PCR amplification using T cell cDNA generated from the same blood sample of line N.

Table 5.2 Summary of all BG cDNA transcripts found in line N (B21)

Line N (B21)	Representative Clones	New Names
NBBGa(97/109)	seq1-1(x24)NB104215	NBBGa-1(24, 2)
	seq1-2(x15)NB118	NBBGa-2(15, 1)
	seq1-3(x5)NB211	NBBGa-3(5, T)
	seq1-4(x4)NB109209	NBBGa-4(4, 2)
	seq1-5(x2)NB124	NBBGa-5(2, 1)
	seq1-6(x2)NB228	NBBGa-6(2, 1)
	seq1-7(x2)NB240	NBBGa-7(2, T)
	seq1-8(x2)NB216	NBBGa-8(2, 1)
	seq1-9(x2)NB217	NBBGa-9(2, 1)
	seq1-10(x1)NB105	NBBGa-10(1, 1)
	seq1-11(x1)NB111	NBBGa-11(1, 1)
	seq1-12(x1)NB140	NBBGa-12(1, 1)
	seq1-13(x1)NB210	NBBGa-13(1, 1)
	seq1-14(x1)NB225	NBBGa-14(1, T)
	seq1-15(x1)NB205	NBBGa-15(1, 1)
	seq1-16(x1)NB201	NBBGa-16(1, 1)
	seq1-17(x1)NB219	NBBGa-17(1, 1)
	seq1-18(x1)NB208	NBBGa-18(1, 1)
	seq1-19(x1)NB229	NBBGa-19(1, 1)
	seq1-20(x1)NB236	NBBGa-20(1, 1)
	seq1-21(x1)NB223	NBBGa-21(1, 1)
	seq1-22(x27)NB320	NBBGa-22(27, V)
NBBGf(12/109)	seq2-1(x1)NB230	NBBGf-1(1, 1)
	seq2-2(x1)NB214	NBBGf-2(1, 1)
	seq2-3(x1)NB312	NBBGf-3(10, V)

Note: Left column: the summary of the BG genes found in line N with the numbers of how many clones were found for this gene out of a total of how many clones were found from line N; the left column being coloured indicated the gene was found in at least two independent PCRs, otherwise the column keeps white background. Middle column: the original record of representative clones that shows a particular BG cDNA sequence with the numbers showing how many clones were found with such sequence. Right column: the final names representing the particular BG cDNA sequence on the left. The final names follow the convention: ‘N’ for line N, ‘B’ for B cells, ‘BG’ and a letter representing the exon 2 sequence, a dash and then a number representing the alternative splicing variant with ‘1’ being the most frequently detected clone (and ‘2’ the second most frequently detected clone, and so forth); the first number in parentheses indicates the number of clones found for a particular exon 2 sequence, and the second number or letter indicates from how many independent PCRs these clones were found with ‘1’ for one PCR, ‘2’ for two PCRs, ‘T’ for one PCR in B cells and one PCR in T cells, ‘V’ for one PCR but only tested for the SS-TM region using SS-TM primers.

5.3.2.2 Three BG genes were found in line P2a (B19)

Three BG genes, P2aBBGa, P2aBBGb and P2aBBGc were found from line P2a (B19) with P2aBBGc dominantly expressed. During the two independent PCRs using HU primers, P2aBBGc was dominantly expressed with only a few clones showing P2aBBGa and P2aBBGb in the first PCR, while in the second PCR, only P2aBBGc was found. To evaluate this result, a third PCR using SS-TM was performed with only P2aBBGc showing up again. As summarized in table 5.3, there are 15 different cDNA sequences found for P2aBBGc but most were found only once in one PCR. It is quite surprising that only one of these transcripts (P2aBBGc-5) appeared in both PCRs using HU primers, with only one clone found in each PCR. Two clones were found once in this project but were also found in one PCR of previous T cell project.

Table 5.3 Summary of all BG cDNA transcripts found in line P2a (B19)

Line P2a (B19)	Representative Clones	New Names
P2aBBGa(1/80)	seq1(x1)PB103	P2aBBGa(1, 1)
P2aBBGb(4/80)	seq2(x4)PB136	P2aBBGb(4, 1)
P2aBBGc(75/80)	seq3-1(x2)PB147	P2aBBGc-1(13, 1st)
	seq3-2(x2)PB236	P2aBBGc-2(13, 1)
	seq3-3(x1)PB228	P2aBBGc-3(8, 2nd)
	seq3-4(x1)PB218	P2aBBGc-4(8, T)
	seq3-5(x1)PB108230	P2aBBGc-5(2, 2)
	seq3-6(x1)PB172	P2aBBGc-6(2, 1)
	seq3-7(x1)PB111	P2aBBGc-7(1, 1)
	seq3-8(x1)PB131	P2aBBGc-8(1, 1)
	seq3-9(x1)PB104	P2aBBGc-9(1, 1)
	seq3-10(x1)PB121	P2aBBGc-10(1, 1)
	seq3-11(x1)PB110	P2aBBGc-11(1, T)
	seq3-12(x1)PB107	P2aBBGc-12(1, 1)
	seq3-13(x1)PB232	P2aBBGc-13(1, 1)
	seq3-14(x1)PB214	P2aBBGc-14(1, 1)
	seq3-15(x1)PB233	P2aBBGc-15(1, 1)
	seq3-16(x20)PB310	P2aBBGc-16(20, V)

Note: same as legend to table 5.2 except the '1st' and '2nd' in P2aBBGc-1(13, 1st) and P2aBBGc-3(8, 2nd) stand for the clones were found from the 1st time PCR and 2nd time PCR, respectively. The reason of specifying from which PCR the two transcripts were found is to help understand such transcripts (encoding soluble BG proteins) were not found in the same PCR, which will be discussed in detail in later sections.

5.3.2.3 Five BG genes were found in line 15I (B15)

Five BG genes, 15iBBGa, 15iBBGb, 15iBBGc, 15iBBGd and 15iBBGe, were obtained in line 15I (B15) in two independent PCRs using HU primers, with 15iBBGa dominantly expressed. 15iBBGa and 15iBBGb were found in both PCRs, and the other genes were only found in one PCR with a few clones. In total, there were 17 different cDNA sequences found for 15iBBGa, with some sequences appearing in both PCRs, some only in one PCR, and some in one PCR here but also in previous T cell project (Table 5.4).

Table 5.4 Summary of all BG cDNA transcripts found in line 15I (B15)

Line 15I (B15)	Representative Clones	New Names
15iBBGa(35/46)	seq1-1(x8)L15B110209	15iBBGa-1(8, 2)
	seq1-2(x5)L15B207	15iBBGa-2(5, T)
	seq1-3(x3)L15B118227	15iBBGa-3(3, 2)
	seq1-4(x3)L15B113	15iBBGa-4(3, 1)
	seq1-5(x2)L15B120	15iBBGa-5(2, 1)
	seq1-6(x2)L15B215	15iBBGa-6(2, 1)
	seq1-7(x2)L15B218	15iBBGa-7(2, 1)
	seq1-8(x1)L15B222	15iBBGa-8(1, 1)
	seq1-9(x1)L15B124	15iBBGa-9(1, 1)
	seq1-10(x1)L15B123	15iBBGa-10(1, 1)
	seq1-11(x1)L15B119	15iBBGa-11(1, 1)
	seq1-12(x1)L15B127	15iBBGa-12(1, 1)
	seq1-13(x1)L15B228	15iBBGa-13(1, 1)
	seq1-14(x1)L15B216	15iBBGa-14(1, 1)
	seq1-15(x1)L15B220	15iBBGa-15(1, 1)
	seq1-16(x1)L15B240	15iBBGa-16(1, T)
	seq1-17(x1)L15B232	15iBBGa-17(1, 1)
15iBBGb(7/46)	seq2-1(x2)L15B125214	15iBBGb-1(2, 2)
	seq2-2(x2)L15B205	15iBBGb-2(2, 1)
	seq2-3(x1)L15B128	15iBBGb-3(1, 1)
	seq2-4(x1)L15B112	15iBBGb-4(1, T)
	seq2-5(x1)L15B231	15iBBGb-5(1, 1)
15iBBGc(2/46)	seq3(x2)L15B229	15iBBGc(2, 1)
15iBBGd(1/46)	seq4(x1)L15B130	15iBBGd(1, 1)
15iBBGe(1/46)	seq5(x1)L15B135	15iBBGe(1, 1)

Note: same as legend to table 5.2.

5.3.2.4 Eight BG genes were found in line 6_1 (B2)

Eight BG genes, 6BBGa, 6BBGb, 6BBGc, 6BBGd, 6BBGf, 6BBGg, 6BBGh and 6BBGi, were obtained from line 6_1 (B2) using HU primers. Three genes, 6BBGa, 6BBGb and 6BBGf were found in both PCRs, while the other genes were only found in one PCR (Table 5.5). Unlike the other three lines, the most highly expressed BG gene found in line 6_1, 6BBGb, was not as dominantly expressed as other BG genes. Also, unlike the dominantly expressed BG genes from the other three chicken lines, 6BBGb has fewer alternatively spliced transcripts. Take line N as a comparison, there are 21 different transcripts found for NBBGa (table 5.2), while only three transcripts were found for 6BBGb from line 6_1 (table 5.5).

By comparing the nearly full-length conceptual transcripts (that is, exons without introns) of all the BG genes found so far, including all the 14 BG genes found in B12 haplotype, six BG genes (6BBGa, 6BBGb, 6BBGc, 6BBGd, 6BBGf, and 6BBGg) found in line 6_1 (B2) are identical to those previously found in B12 (Appendix I). As discussed already in previous T cell project (section 4.3.4.4 chapter 4), B2 and B12 haplotypes share some serological identity which is most likely due to the identical BG genes.

Table 5.5 Summary of all BG cDNA transcripts found in line 6_1 (B2)

Line 6_1 (B2)	Representative Clones	New Names
6BBGa(5/63)	seq1-1(x3)L6B177	6BBGa-1(3, 1)
	seq1-2(x1)L6B190	6BBGa-2(1, 1)
	seq1-3(x1)L6B206	6BBGa-3(1, 1)
6BBGb(27/63)	seq2-1(x17)L6B181203	6BBGb-1(17, 2)
	seq2-2(x7)L6B189	6BBGb-2(7, 1)
	seq2-3(x3)L6B111	6BBGb-3(3, 1)
6BBGc(6/63)	seq3(x6)L6B158	6BBGc(6, 1)
6BBGd(5/63)	seq4(x5)L6B218	6BBGd(5, 1)
6BBGf(12/63)	seq5-1(x7)L6B220	6BBGf-1(7, 1)
	seq5-2(x4)L6B166	6BBGf-2(4, 1)
	seq5-3(x1)L6B219	6BBGf-3(1, 1)
6BBGg(6/63)	seq6-1(x3)L6B152	6BBGg-1(3, 1)
	seq6-2(x3)L6B122	6BBGg-2(3, 1)
6BBGh(1/63)	seq7(x1)L6B211	6BBGh(1, 1)
6BBGi(1/63)	seq8(x1)L6B233	6BBGi(1, 1)

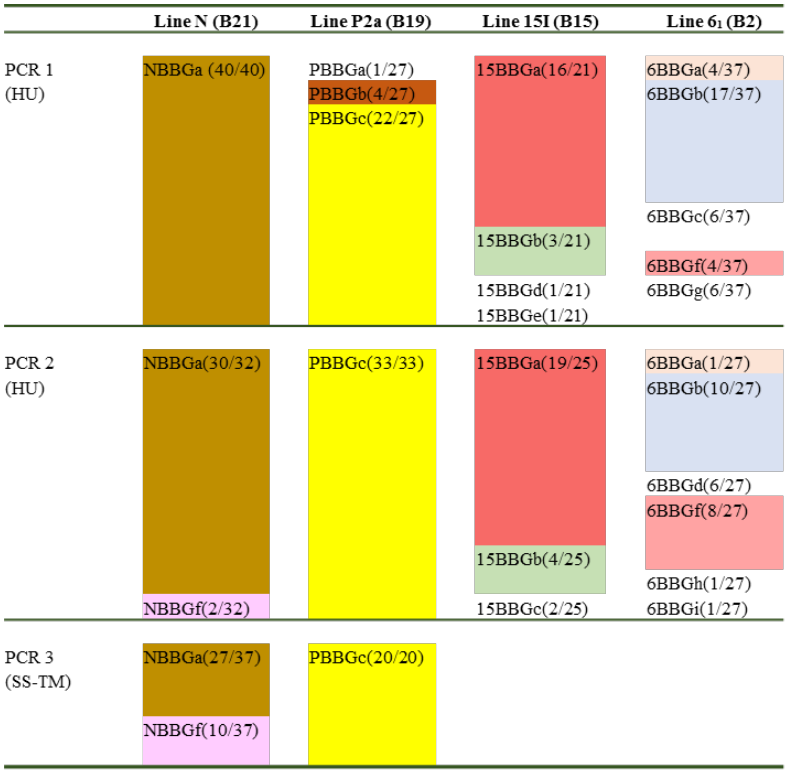
Note: same as legend to table 5.2.

This page intentionally left blank

5.3.2.5 Summary: 18 BG genes were found in four different chicken lines, each with one BG gene dominantly expressed

As shown in figure 5.1, eighteen different BG genes (based on exon 2 sequences) were detected from peripheral B cells of four different chicken lines. None of these genes is identical, but six BG genes from line 6_1 (B2) are also found in line CB (B12) (Appendix I). For each line, two independent PCRs were performed using HU primers, and an extra PCR using SS-TM primers was performed for line N and line P2a in order to evaluate the HU PCR results.

Generally, there was one dominantly expressed BG gene found in each line, with three genes (NBBGa, P2aBBGc and 15iBBGa) dominantly expressed in the three lines. These dominantly expressed BG genes were found with many different alternative splicing transcripts. On the contrary, the most dominantly expressed BG gene in line 6_1 (6BBGb) was not highly expressed, nor found with many different transcripts.



Line CB (B12)	Line N (B21)	Line P2a (B19)	Line 15I (B15)	Line 6₁ (B2)
BG7(16/40)	NBBGa(97/109)	P2aBBGa(1/80)	15IBBGa(35/46)	6BBGa(5/63)
BG8(2/40)	NBBGf(12/109)	P2aBBGb(4/80)	15IBBGb(7/46)	6BBGb(27/63)
BG9(18/40)		P2aBBGc(75/80)	15IBBGc(4/46)	6BBGc(6/63)
BG12(1/40)			15IBBGd(1/46)	6BBGd(5/63)
BG13(3/40)			15IBBGe(1/46)	6BBGf(12/63)
				6BBGg(6/63)
				6BBGh(1/63)
				6BBGi(1/63)

Figure 5.1 Overall results for the number of BG genes amplified from B cells of four chicken lines. Top panel, independent amplifications from the four chicken lines; HU, haemopoietic forward and 'universal' reverse primers to give nearly full-length sequences; SS-TM, signal sequence forward and transmembrane reverse primers to give SS, extracellular Ig-V domain and TM regions. Different colours indicate different exon 2 sequences, except those sequences that are only found in one PCR reaction. Names follow the convention: abbreviated line name, 'B' for B cells, 'BG', and a letter representing the exon 2 sequence; numbers in parentheses indicate the number of clones found for a particular exon 2 sequence out of the total number for the particular PCR reaction. Bottom panel, the total results for four chicken lines from this project and for the line CB (B12) from Salomonsen *et al*., 2014.

5.3.3 Alternative splicing was observed in most BG transcripts, with cDNA sequences potentially encoding soluble BG proteins detected

Alternative splicing was observed for most BG transcripts found in B cells. Previously in the T cell project (chapter 4), it was found that the alternative splicing in BG cDNA sequences was mainly due to the intron retention. In order to illustrate the intron-exon organization, all the BG cDNA transcripts found from the same line, except those sequences with obvious PCR errors, were aligned against the genomic sequence of BG8 from B12 haplotype (Appendix J, Figure 5.2).

The alignments show that two major factors contribute to the alternative splicing. The most frequent one is intron retention which introduces an early stop codon in protein translation. The retained introns are mainly found in the cytoplasmic tail regions of all four chicken lines, resulting in truncated cytoplasmic tails. A few sequences with retained introns between Ig-V domains and transmembrane regions were also observed but only in three chicken lines, line P2a, 15I and 6_1; these transcripts may encode soluble BG proteins containing only Ig-V domains. The other factor is the usage of alternative splice acceptor sites (Chattaway, 2013) in the cytoplasmic tail regions. For instance, NBBGa-1(24, 2) and NBBGa-2(15, 1) are almost identical in intron-exon organization structure and are same in sequence lengths, 1772 bp (Figure 5.2); however, the latter sequence uses alternative splice acceptor sites twice in the cytoplasmic tail region, with one introducing an insertion of three nucleotides and the other one introducing a deletion of three nucleotides (Appendix J1).

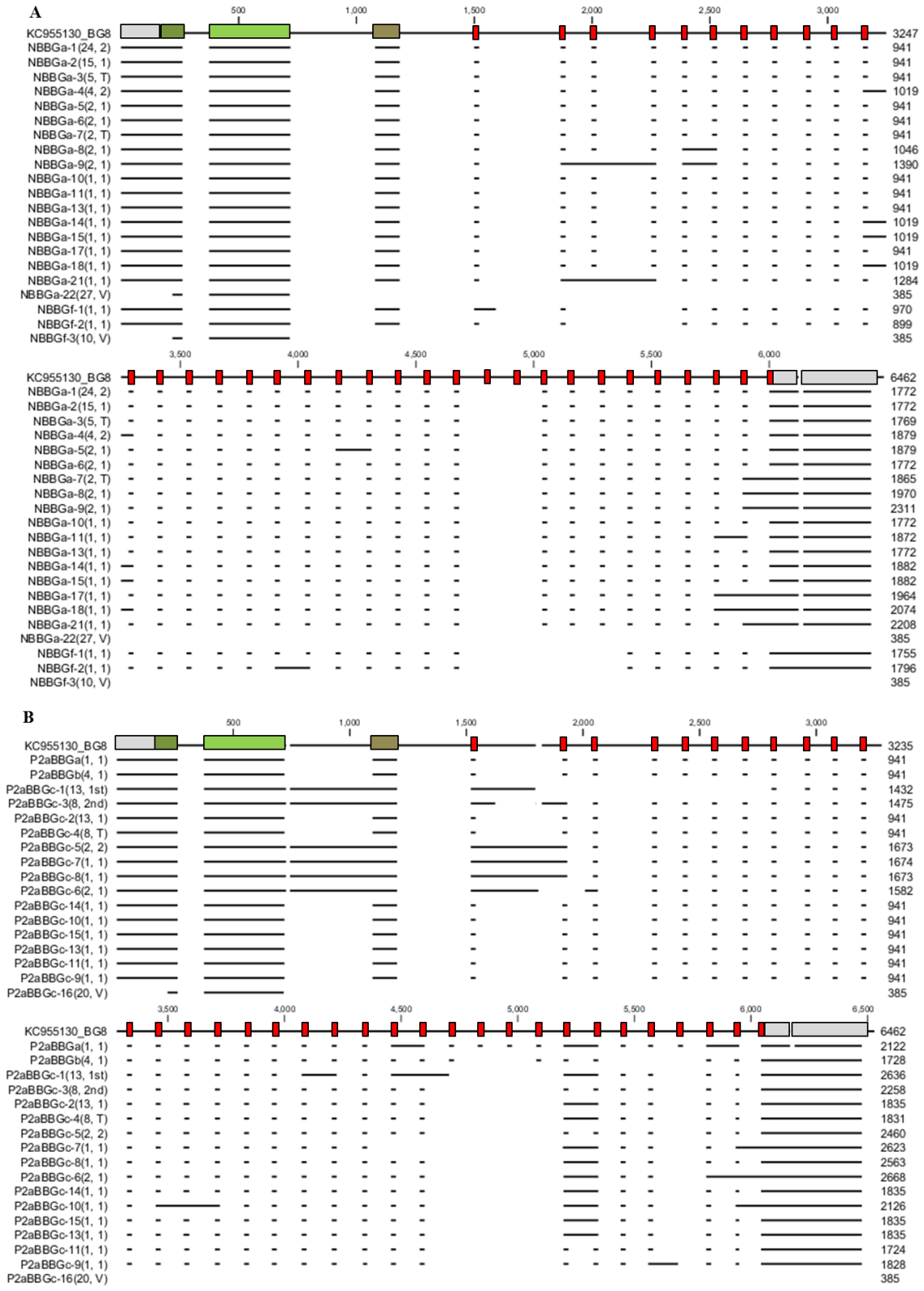


Figure legend is shown on the page following the figure

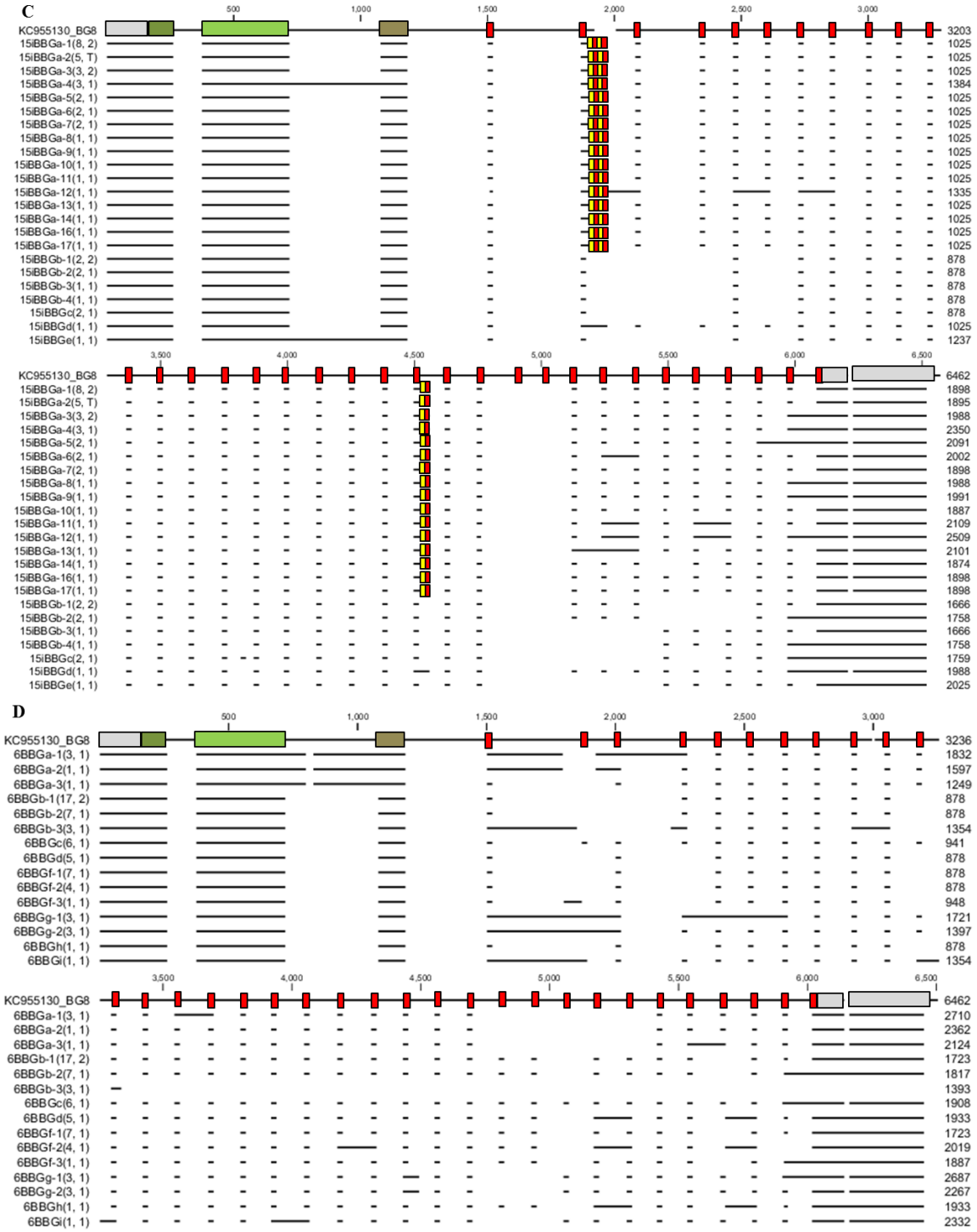


Figure legend is shown on the page following the figure

Figure 5.2 Representation of the intron-exon structure of the BG genes inferred from all the alternatively-expressed transcripts found from B cells of four chicken lines. The top line represents the genomic sequence of BG8 from line CB (B12) with boxes indicating exons: grey for 5' UTR and 3' UTR, dark green for signal sequence, light green for Ig-V domain, brown for transmembrane, red for cytoplasmic tail. The actual mRNA transcripts from line N (A), line P2a (B), line 15I (C) and line 6_1 (D) are indicated as horizontal lines. In figure C, the alternating red and yellow boxes indicate the additional heptad repeats found for this gene, compared to BG8 (B12). Names of the transcripts follow the convention: abbreviated line name, 'B' for B cells, 'BG' and the letter representing a particular exon 2 sequence, a dash and then a number representing the alternative splicing variant with '1' being the most frequently detected clone (and '2' the second most frequently detected clone, and so forth). Numbers in parentheses indicate the number of clones found for a particular full sequence, followed by the number of independent PCRs or a letter in which the sequence was identified (1, found in one PCR; 2, found in 2 PCRs; T, found in one PCR in this project and one using T cell cDNA in previous T cell project).

5.3.4 All BG genes found in B cells have haemopoietic 5' UTRs and most fall into BG8-9-12 clade and BG5-7-11 clade

In order to understand the relationships of the 18 BG genes found in B cells of four chicken lines, together with the 14 BG genes in line CB (B12), the nearly full-length conceptual cDNA sequences (that is, exons without introns) of all 32 BG genes were analyzed by phylogenetic trees of nucleotide sequences using Neighbour Joining method with bootstrap (1000). Overall, all 18 BG genes found in B cells have haemopoietic BG features (haemopoietic 5' UTRs) with either type 1 or type 2 cytoplasmic tails. They are all grouped closely to either BG8-9-12 clade or BG5-7-11 clade, except one gene from line N, NBBGf (Figure 5.3). NBBGf has the haemopoietic 5' UTR but is grouped closely to the tissue BG gene BG6 in other regions except the 3' UTR region (Figure 5.4).

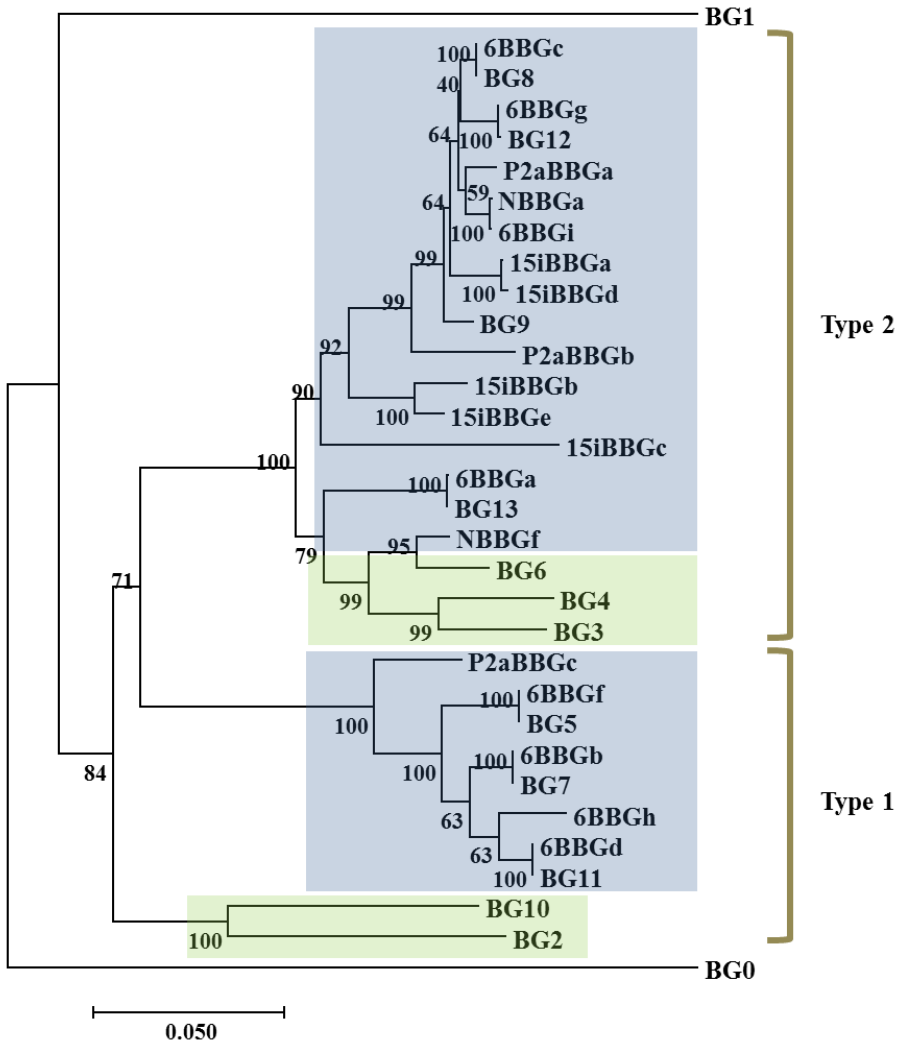


Figure 5.3 Phylogenetic tree built on full-length conceptual transcripts of all 18 BG genes found in B cells from four chicken lines and all 14 BG genes from B12 haplotype. Names of the transcripts follow the convention: abbreviated line name, 'B' for B cells, 'BG' and the letter 'a' representing a particular exon 2 sequence. Names of the genes follow the convention 'BG' and the number of the gene locus from the B12 haplotype. Indicated by colour are those clades with 5' ends of haemopoietic (blue) and tissue (green), and by brackets for 3' ends of type 1 and type 2. Branch lengths are scaled by genetic distance, and percentage bootstrap values out of 100 are indicated at the nodes.

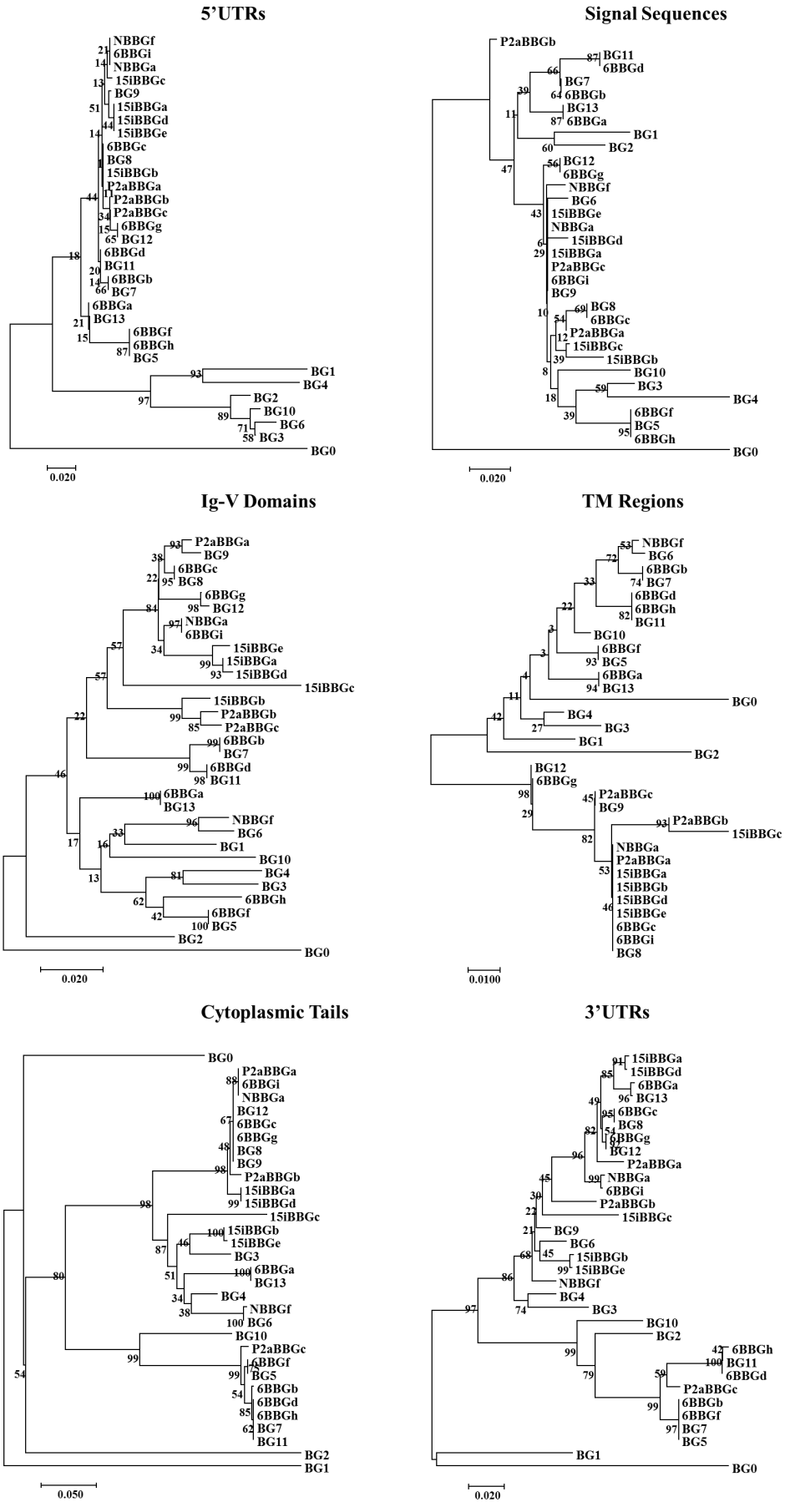


Figure legend is shown on the page after the figure

Figure 5.4 Phylogenetic trees of nucleotide sequences for different regions of the full-length conceptual transcripts for all 18 BG genes found in B cells from four chicken lines and 14 BG genes from B12 haplotype. Trees include 5' UTR from exon 1, signal sequence from exon 1, exon 2 (two nucleotides from the signal sequence, and then the nucleotides encoding the Ig-V domain), exon 3 (transmembrane region), the exons corresponding to the cytoplasmic tail (excluding any nucleotides in the final exon that encode amino acids), and the final exon (which is exactly the 3' UTR in some sequences, but for which the first nucleotides encode the last amino acids of the cytoplasmic tail in most sequences). Other details are as in the legend to figure 5.3.

The 5’ UTRs of all 18 BG genes found in this project fall into the same group with the well characterized ‘haemopoietic BGs’ of B12 haplotype, illustrating that they are all haemopoietic BG genes. In fact, all the 18 BG genes were obtained from B cells, which should have the haemopoietic BG gene character. Moreover, the primers used for BG gene amplification was designed specifically for haemopoietic BG genes. Salomonsen et al. have proposed that all the ‘haemopoietic BGs’ in the B12 haplotype are descended from a single ancestor (Salomonsen *et al.*, 2014); and this assumption is supported by our alignments that all ‘haemopoietic BGs’ from four chicken lines also have the same large deletion in the 5’ UTR region compared to ‘tissue BGs’ (Figure 5.5).

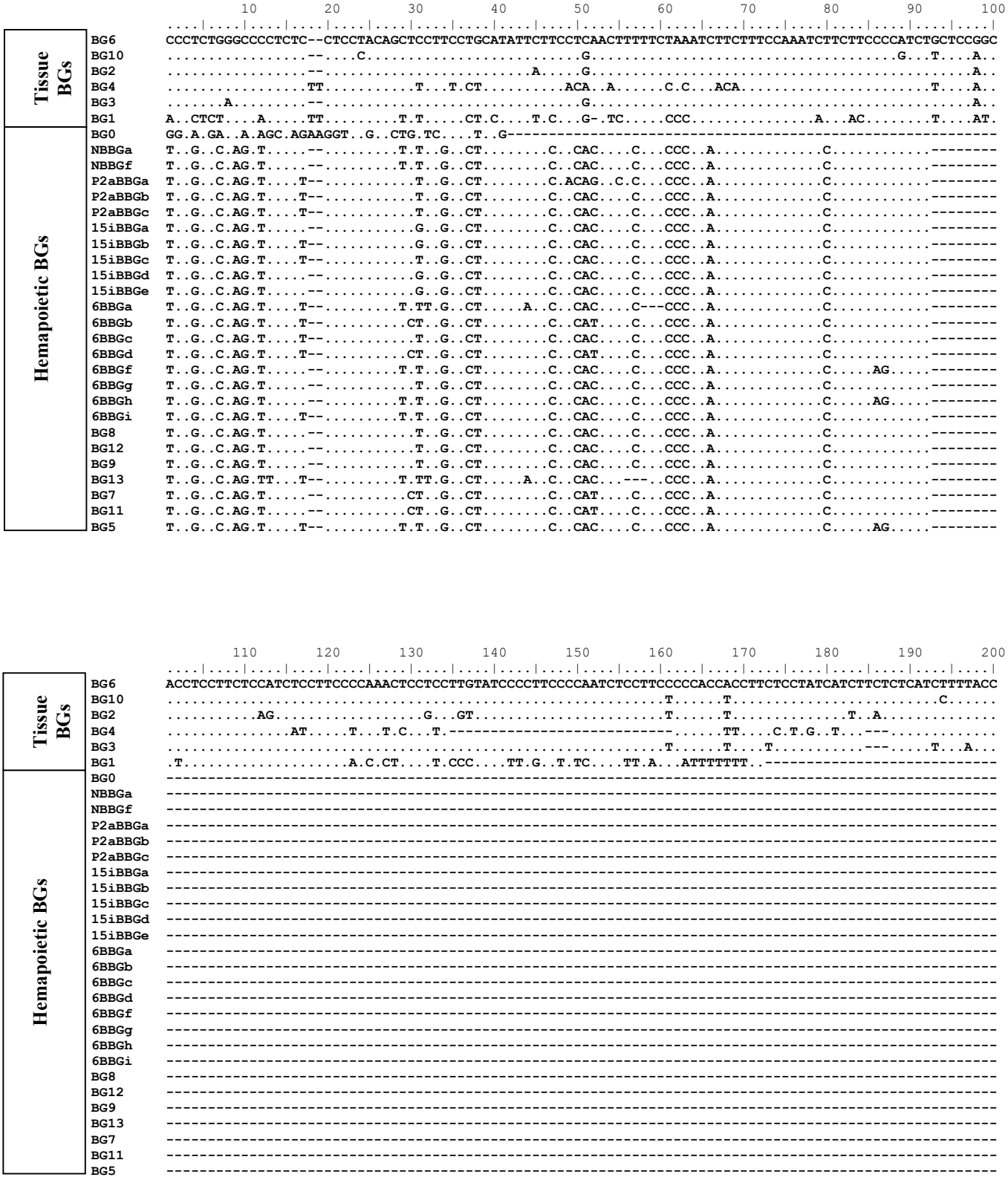
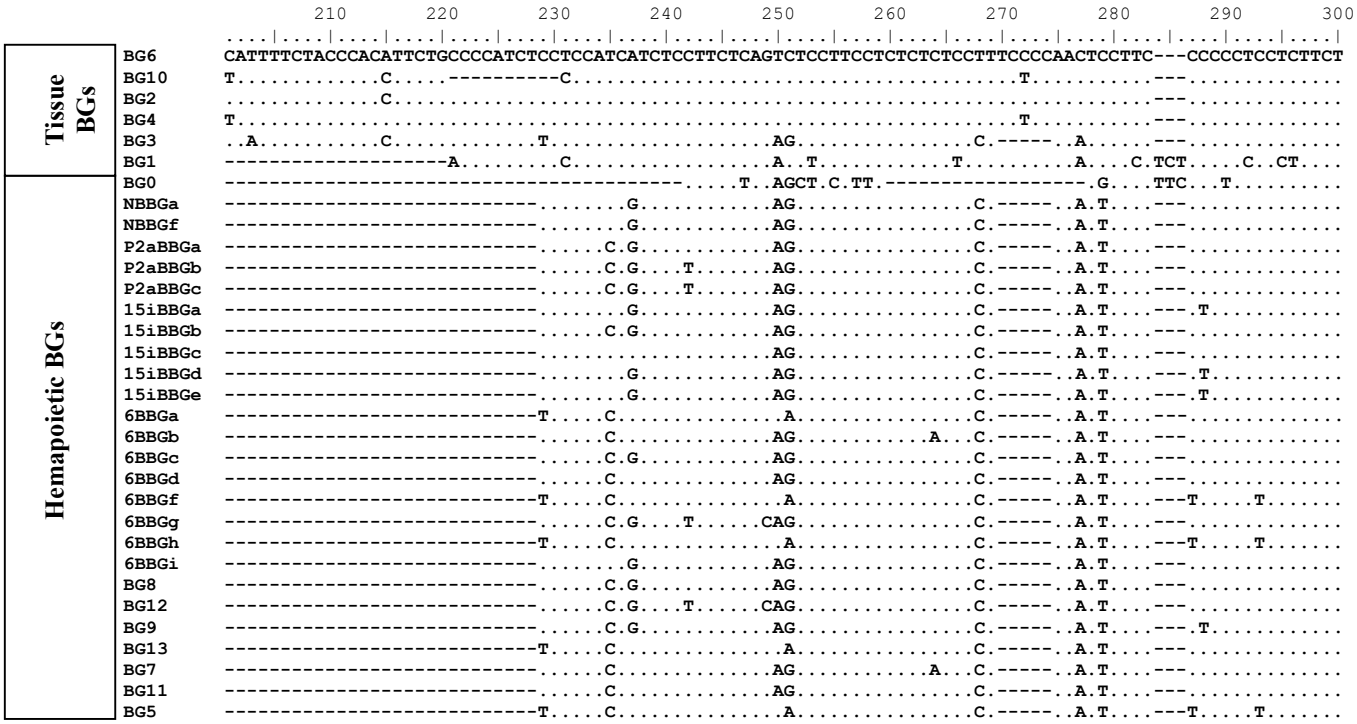


Figure legend is shown on the page after the figure



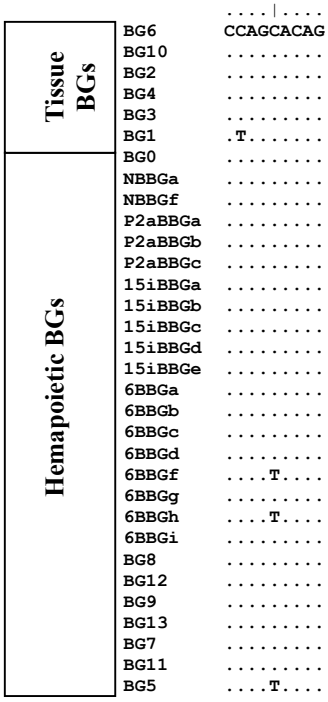


Figure 5.5 The alignment of 5' UTRs between 18 BG genes found in B cells of four chicken lines and all 14 BG genes from B12 haplotype. The alignment were made with the 5' UTR sequences of all 18 BG genes found from B cells of four different chicken lines [line N (B21), line P2a (B19), line 15I (B15) and line 61 (B2)] and all 14 BG genes found in line CB (B12). Compared to the tissue BGs that have been well characterized in previous research, there is a large gap in the 5' UTR sequences of haemopoietic BGs (Salomonsen et al., 2014); the alignment shows that all the 18 BG genes found in this project have haemopoietic 5' UTRs. Letters indicate nucleotides, dots indicate identities with BG6 sequence on the top, dashes indicate no sequence present compared to one or more of the other sequences.

The length of signal sequences are too short (99 bp) to produce a reliable phylogenetic tree; also, from the tree there is no obvious pattern to help understand the relationships of these genes.

The phylogenetic tree built on Ig-V domains shows great polymorphism of the BG genes, but with most genes found in this project distributed into the three groups identified already in the V domain dendrogram built on B12 BG genes (Salomonsen *et al.*, 2014) (Figure 5.4). The three major Ig-V domain groups observed by Salomonsen are: BG3-4-5, BG7-11 and BG8-9-12. Most of the 18 BG genes fall into the BG8-9-12 group with a few in the other two groups. One exception is these two genes, 6BBGa and NBBGf, with 6BBGa found identical to BG13, and NBBGf very similar to BG6.

The transmembrane regions are too short (105 bp) to reflect a very reliable phylogeny, especially when many genes have virtually identical sequences. However, all 18 BG genes were separated into the two groups already identified by Salomonsen (Salomonsen *et al.*, 2014), with 12 BG genes in the BG8-9-12 group and the other six BG genes in the other group, which includes all the other BG genes found in B12 haplotype (Figure 5.4).

The phylogenetic tree built on cytoplasmic tails also fits with the previous study. All 18 BG genes fall into two major types, type 1 and type 2 (Salomonsen *et al.*, 2014), and under type 2, all BG genes were divided into two subtypes, type 2a and type 2b (Chattaway, 2013; Chen *et al.*, 2018) (Figure 5.4).

Two types of 3' UTRs were identified from the phylogenetic tree, and all 18 BG genes fall into two groups. This finding again is consistent with the two types of 3' UTR from previous work (Salomonsen *et al.*, 2014), which also provides evidence that the HU primer pair could amplify haemopoietic BG genes from either type 1 or type 2 3' UTRs.

To summarize, all 18 BG genes found in B cells from four chicken lines have the haemopoietic 5' UTRs with either type 1 or type 2 3' UTRs. Phylogenetic studies based on whole cDNA sequence of the 18 BG genes as well as the well-characterized 14 BG genes from B12 haplotype showed that all the 18 BG genes either belong to BG8-9-12 clade, or BG5-7-11 clade, except one gene (NBBGf) closely related to a tissue BG gene (BG6). However, looking at separate regions, some BG genes are clustered together in one region but separated into different groups in other regions, which makes BG genes look like hybrid

genes. This suggests that recombination play a role during BG evolution (Salomonsen *et al.*, 2014).

5.3.5 Dominantly expressed BG genes in B cells from the four chicken lines are distributed to two BG clades except for one gene whose Ig-V domain does not show a clear pattern

In previous research on the B12 haplotype, two almost equally expressed BG genes were found in B cells (BG7 and BG9) with a few clones for three other BG genes (BG8, 12 and 13). BG7 and BG9 belong to two different BG clades, the BG7-11 clade and the BG8-9-12 clade, respectively. BG8-9-12 clade and BG7-11 clade show many differences throughout the sequences except in the 5' UTRs, and long distances in the phylogenetic trees made with all 14 BG genes of B12 haplotype (Salomonsen *et al.*, 2014). In this project, we expected that there would be two kinds of BG genes dominantly expressed in B cells of the other haplotypes as well, with one gene clustered into the BG8-9-12 clade and the other one going into the BG7-11 clade. To our surprise, as shown in figure 5.1 already, only one dominantly expressed BG gene was found from each of the three chicken lines, NBBGa from line N, P2aBBGc from line P2a, and 15iBBGa from line 15I. One dominantly expressed BG gene (6BBGb) was found from line 6_1 with a few sub-dominantly expressed BG genes (6BBGf, 6BBGc, 6BBGg, 6BBGa and 6BBGd).

Nucleotide sequence alignments and phylogenetic studies of all four dominantly expressed BG genes (NBBGa, P2aBBGc, 15iBBGa and 6BBGb) together with the 14 BG genes from B12 haplotype reveal that two BG genes (NBBGa and 15iBBGa) belong to the BG8-9-12 clade, one BG gene (6BBGb) belongs to the BG7-11 clade, and the other BG gene (P2aBBGc) doesn't cluster with either group. As shown in figure 5.6 and figure 5.7, P2aBBGc looks like a hybrid gene with a haemopoietic 5' UTR, an Ig-V domain with another two BG genes (P2aBBGb and 15iBBGb) forming a separate group, the same type of transmembrane as BG8-9-12 clade, and a typical type 1 cytoplasmic tail and 3' UTR which groups with BG7-11 clade. Because the length of cytoplasmic tail together with 3' UTR is longer than the other regions, the phylogenetic tree built on whole sequence shows P2aBBGc grouped with BG7-11 clade. Therefore, the conclusion from current results shows that NBBGa, 15iBBGa and BG9 are functional alleles belonging to BG8-9-12 clade, 6BBGb and BG7 are functional alleles which belong to BG7-11 clade, and P2aBBGc clusters with BG7-11 clade in cytoplasmic tail.

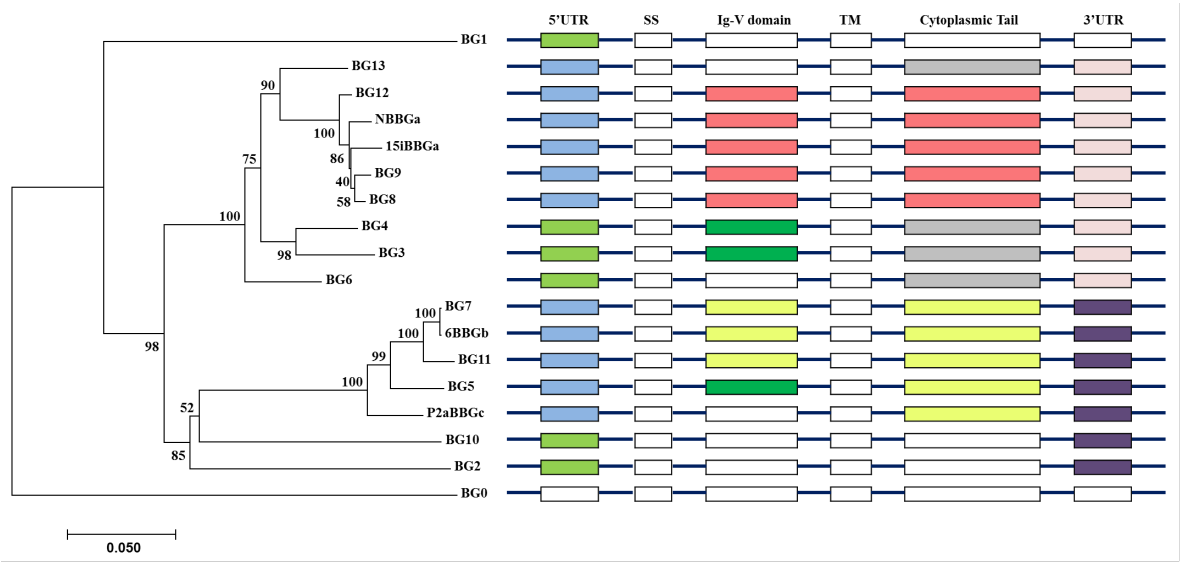


Figure 5.6 Overview of the phylogenetic relationship of all the four dominantly expressed BG genes found in B cells from four chicken lines with the 14 BG genes from B12 haplotype. Phylogenetic tree on the left was built on the whole conceptual cDNA sequences. On the right side, the same colours under separated regions indicate the same clusters when phylogenetic tree was built for that particular region in figure 5.4.

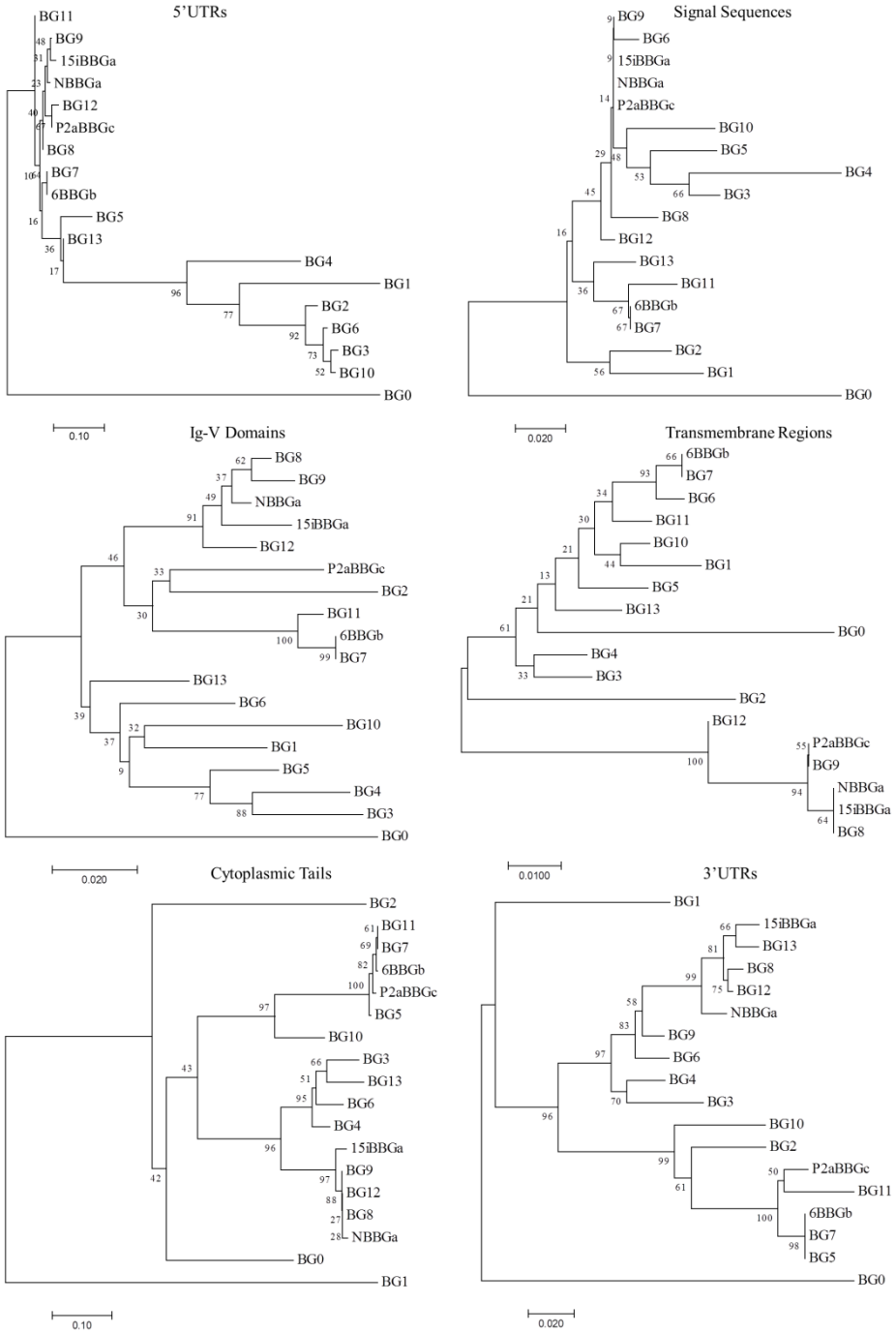


Figure 5.7 Phylogenetic trees of nucleotide sequences for different regions of the full-length conceptual transcripts for dominantly expressed BG genes in B cells from four chicken lines and 14 BG genes from B12 haplotype. Trees include 5' UTR from exon 1, signal sequence from exon 1, exon 2 (two nucleotides from the signal sequence, and then the nucleotides encoding the Ig-V domain), exon 3 (transmembrane region), the exons corresponding to the cytoplasmic tail (excluding any nucleotides in the final exon that encode amino acids), and the final exon (which is exactly the 3' UTR in some sequences, but for which the first nucleotides encode the last amino acids of the cytoplasmic tail in most sequences). Other details are as in the legend to figure 5.3.

5.3.6 BG8-9-12 clade shows evidence of selection for variation in cytoplasmic tail but not in other regions

We expected that the dominantly expressed BG gene in B cells from different haplotypes would have the same function, and thus could be treated as 'functional alleles'. Therefore, comparing these functional alleles by looking at their sequences at both the nucleotide and amino acid levels, as well as the location and potential clustering of the sequence variation, would give us an insight into the features of these sequences, whether there is evidence of selection for variation, which regions of the BG genes are important for function, etc. Thus far, two BG clades, BG8-9-12 clade and BG7-11 clade, have been discovered to distinguish different BG alleles. Only those BG genes clustered within the same clade throughout the whole cDNA sequence analysis are considered in a BG clade.

The BG8-9-12 clade discovered in B cells include three BG genes, BG9 from line CB (B12), NBBGa from line N (B21) and 15iBBGa from line 15I (B15). Both NBBGa and 15iBBGa are the same genes as NTBGa and 15iTBGa found from T cells, respectively. NTBGa and 15iTBGa have been found as the 'functional alleles' together with another three BG genes (P2aTBGa, 6TBGa and BG9) in T cell project, and have been examined for the silent versus replacement changes in their sequences, with the results indicating selection for variation in cytoplasmic tail but not in other regions. Detailed information can be found in section 4.3.6 (chapter 4).

5.3.7 BG7-11 clade from line CB and line 6_1 are identical in sequence; the dominantly expressed BG gene in line P2a has the cytoplasmic tail with typical BG7-11 clade features but showing evidence of variation for selection

The functional BG alleles that clearly belong to BG7-11 clade from B cell project are 6BBGb and BG7, which are identical in sequence (Appendix I). The dominantly expressed BG gene in B cells from line P2a, P2aBBGc, does not fall into either clustered clade, although it has a cytoplasmic tail from the BG7-11 clade.

All BG genes in BG7-11 clade, including P2aBGc, have 29 heptad repeats in the cytoplasmic tails. The cytoplasmic tails of BG genes are composed of amino acid heptad repeats encoded by 21 nucleotide exons (with a few exons of 18 or 24 nucleotides), and it has been well recognized that the numbers of heptad repeats vary among BG genes. The heptad repeat

numbers are the same for the three BG genes above, suggesting functional importance for the BG7-11 clade.

Furthermore, examination of variation in sequence at both nucleotide and amino acid levels between the two BG genes (as 6BBGb is identical to BG7), BG7 and P2aBBGc, revealed a significant finding. Compared to BG7, there are 19 nucleotide differences with only one silent change but 17 replacements, as one codon involves two nucleotide changes in P2aBBGc (Figure 5.8, Figure 5.9). It is striking that virtually all nucleotide changes are nonsynonymous changes, indicating the variation is under selection. Detailed examination of cytoplasmic tail features including the amino acid variation is carried out in later section 5.3.10.

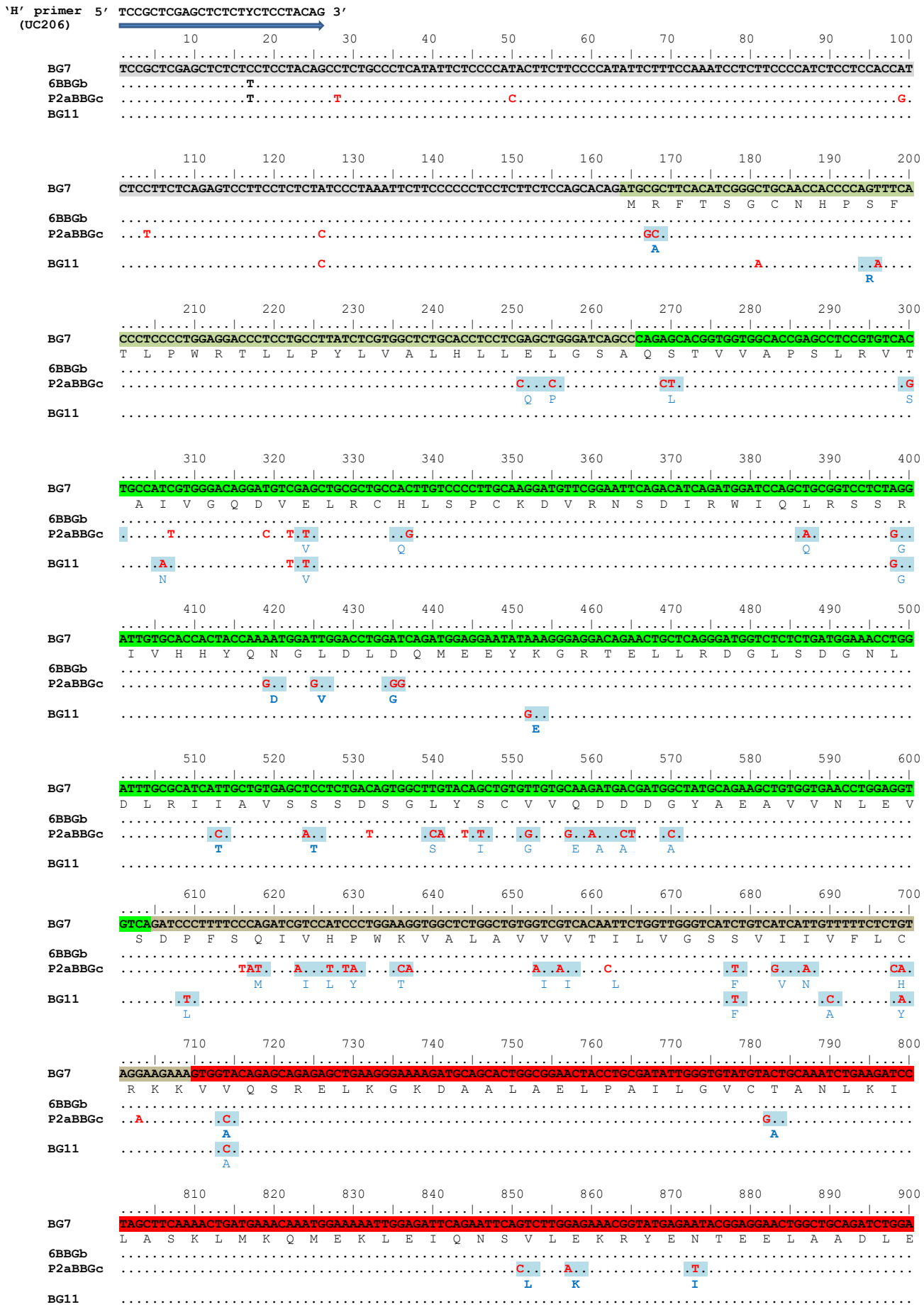


Figure legend is shown on the page following the figure

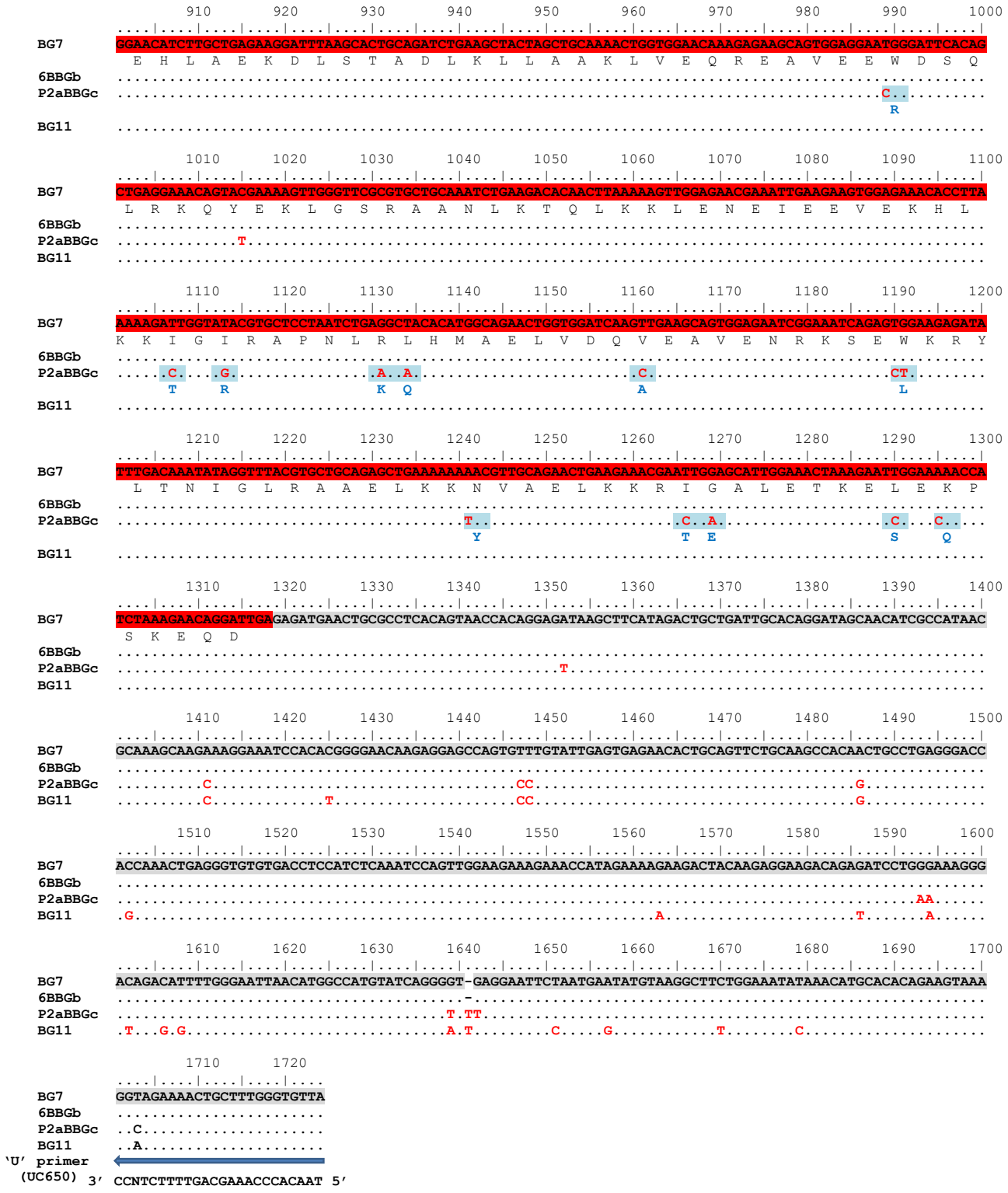


Figure legend is shown on the page following the figure

Figure 5.8 Alignment of nucleotide sequences (along with amino acid positions that differ from the BG7 gene from the B12 haplotype) for portions of the dominantly-expressed BG genes in B cells from two chicken lines, and for the appropriate BG genes of the B12 haplotype. BG7, 6BBGb and P2aBBGc are the dominantly expressed BG genes found from line CB, line 6_1 and line Pa2, respectively. Portions of the sequence analyzed are: signal sequence from exon 1 labeled dark green, exon 2 (mostly Ig-V domain) labeled light green, exon 3 (transmembrane region) labeled brown, and exons corresponding to the cytoplasmic tail (including translated sequences in the last exon) labeled red. Letters indicate nucleotides, dots indicate identities with BG7 sequence, dashes indicate no sequence present compared to one or more of the other sequences. Codons corresponding to nucleotide changes that lead to amino acid changes are boxed. Arrows indicate primers with sequences (so that sequence differences in these positions are not necessarily real).

		S (1)	S (2)	S (3)	N (1)	N (2)	N (3)	N (1+2)	N (1+3)	N (2+3)	N (1+2+3)
Signal Sequence	6BBGb										
	P2aBBGc				1	1		1			
	BG11			1			1				
Ig-V Domain	6BBGb										
	P2aBBGc			5	6	7	1	1		3	
	BG11			1	2	2					
TM Region	6BBGb										
	P2aBBGc	1		2	4	3		3		1	
	BG11					4					
Cytoplasmic Tail	6BBGb										
	P2aBBGc			1	6	10		1			
	BG11					1					

Figure 5.9 Compared with BG7 of the B12 haplotype, the number of silent and replacement changes by codon position for the dominantly expressed BG genes in B cells from two chicken lines, and of the other appropriate BG gene from B12 haplotype. Names of the transcripts follow the convention: abbreviated line name, 'B' for B cells, 'BG' and the letter representing a particular exon 2 sequence. Names of the genes follow the convention 'BG' and the number of the gene locus from the B12 haplotype. Values are based on the alignments in figure 5.8; amino acids from split codons at the edges of the exons are assigned to the exon with two of the three nucleotides of the codon (for instance, last amino acid of the signal sequence is assigned to the Ig-V domain, which in fact starts with glutamine in the mature protein).

5.3.8 The variation in Ig-V domain structure among all the dominantly expressed BG genes found in B cells have the amino acid side chains pointing out

As mentioned above, the Ig-V domains of all the dominantly expressed BG genes found in B cells belong to three groups, the BG8-9-12 clade, the BG7-11 clade, and the group formed by P2aBBGc with another two BG genes (P2aBBGb and 15iBBGb). It has been found that the variation in the Ig-V domain of T cell BG genes in the BG8-9-12 clade is mostly located in the loop region (section 4.3.7 chapter 4). It would be interesting to see the differences among these three groups and where the variable residues are located.

Comparing the Ig-V domains of all the dominantly expressed BG genes from five chicken lines, quite a lot of variations are found along the whole sequence, although there are some conserved regions. For example, regions from position 30 to position 40, from position 58 to position 82 (from strand C'' to strand E), and from position 103 to position 113 are conserved in the amino acid sequence alignments (Figure 5.10). Also there is no change in the cysteines that form the intra-domain disulfide bond (position 23 and position 95 in figure 5.10), or the cysteine located in the equivalent of complementarity determining region 1 (CDR1) that forms a disulfide bond between the two chains of a BG dimer (position 28 in figure 5.10).

The variable amino acids among these dominantly expressed BG genes are located not only in the loop region, but also in the β strands. More strikingly, they all have their side chains pointing to the outside of the domain. As shown in figure 5.10, the variable amino acids are labeled purple and annotated as sticks on both the Ig-V domain structures built for BG8 monomer and BG1 homodimer. Clearly, all the variable amino acids are pointing to the outside, suggesting the BG genes from the three groups may interact with different ligands (or receptors). Alternatively, all the residues pointing inwards towards the centre of the two β sheets are constrained and therefore conserved, whereas the residues pointing outwards are free to change because they have no function.

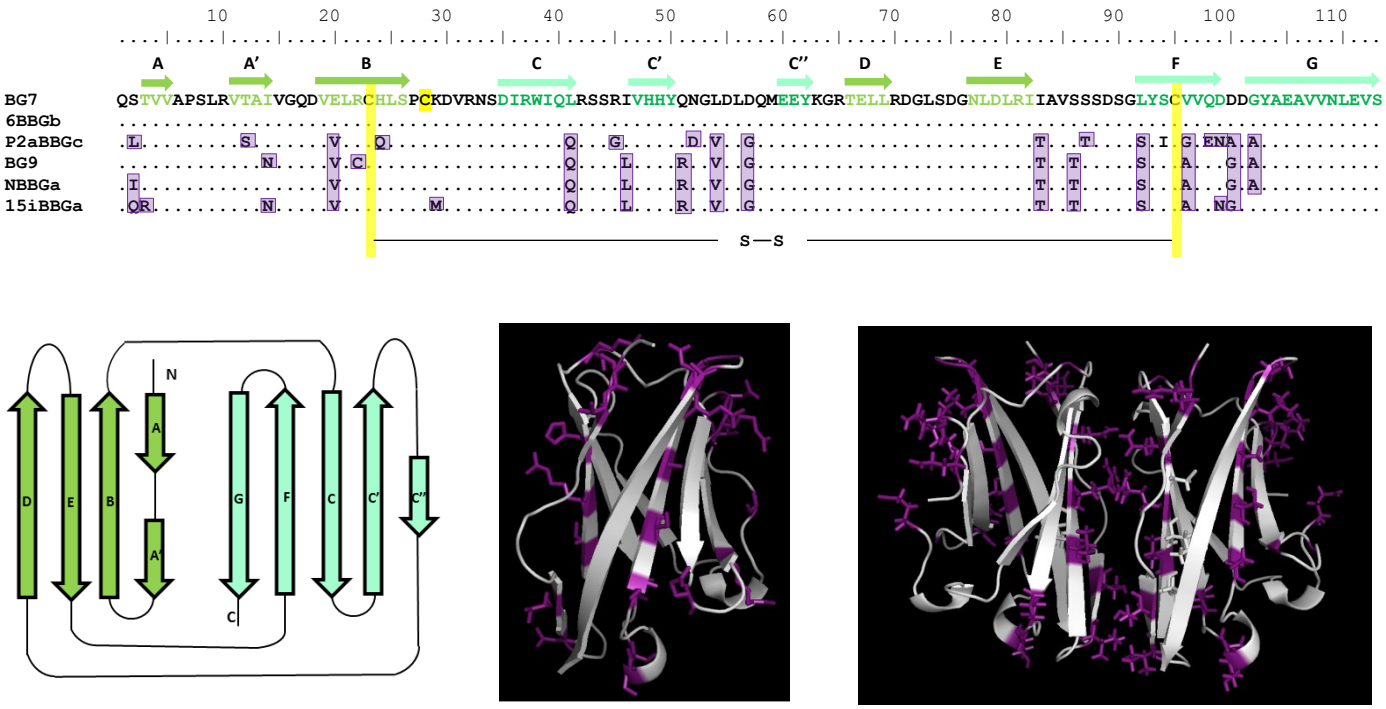


Figure 5.10 Alignment of amino acid sequences for the Ig-V domains of the dominantly-expressed genes in B cells from four chicken lines, and the two almost equally dominantly expressed BG genes in B cells of B12 haplotype from Salomonsen et al 2014, along with structural models of the Ig-V domains with the location of variation (labeled in purple) compared to the BG7 sequence of the B12 haplotype indicated. The Ig-V structure model on the left is BG8, while the dimer structure on the right is BG1. In the top panel, letters indicate amino acids by single letter code, dots indicate identities with BG7 sequence, residues that differ from BG7 are boxed in purple for the other genes. The β-strands of the V region are indicated by arrows in the top panel, and are coloured dark green for one face of the domain and light green for the other face. The same colour scheme is used for the three panels below.

5.3.9 Helical wheel structures of cytoplasmic tails show striking differences between the two types of cytoplasmic tails from dominantly expressed BG genes of five different chicken lines

As discussed previously in T cell project, the BG cytoplasmic tails are encoded by many 21 nucleotide exons, and the 4th and 7th codons in one exon correspond to the first and fourth amino acids (also called the *a* and *d* positions) of the true heptad repeat (section 4.3.8 chapter 4). The α-helix structures of the cytoplasmic tails of all the dominantly expressed BG genes are therefore organized following the order of 4th codon (*a* position), 7th (*d*), 1st (*e*), 5th (*b*), 2nd (*f*), 6th (*c*), 3rd (*g*), as presented in figure 5.11. In such a structure, two α-helical chains could form a dimer, called a coiled-coil (Kaufman *et al.*, 1989; Kaufman *et al.*, 1990), with the *a* and *d* positions acting as the interface between the two chains, together with some contribution by the neighbouring amino acids (*e* and *g*) (Aronsson *et al.*, 2015). However, it is still unclear from these pictures that whether they are homo-dimers or hetero-dimers. From the comparison of cytoplasmic tail structures of all dominantly expressed BG transcripts, two different types of cytoplasmic tails are clearly observed. One type includes three BG genes, BG9, NBBGa and 15iBBGa, which all come from the BG8-9-12 clade. The other type includes the rest of the BG genes, BG7, 6BBGb and P2aBBGc, with BG7 and 6BBGb from the BG7-11 clade. These two types of tails differ from each other hugely but remain conserved within the same clade. The different and common features are summarized as bellows.

First, one obvious difference is the location of conserved cysteines. In BG7-11 clade including P2aBBGc, the cysteine is encoded by the 3rd codon of the 4th heptad repeat which is close to the transmembrane region, while in BG8-9-12 clade the cysteine is encoded by the 4th codon of the heptad repeat that is close to the C-terminus.

Second, the numbers of heptad repeats are the same within BG7-11 clade (including P2aBBGc) but vary in the BG8-9-12 clade. In BG8-9-12 clade, there are two insertions in 15iBBGa compared to the other two genes (BG9 and NBBGa). The first insertion is four heptad repeats after the second heptad repeat. The second insertion is five heptad repeats if compared to BG9 or two heptad repeats if compared to NBBGa, starting from the third heptad repeat after the conserved cysteine.

Third, the amino acid charges in the five positions other than *a* and *d* positions are clustered

into different patches in BG8-9-12 clade and BG7-11 clade. Overall, the cytoplasmic tails of BG8-9-12 clade are acidic (158 D+E versus 132 K+R), while the tails of BG7-11 clade are more neutral in charge (113 D+E versus 116 K+R).

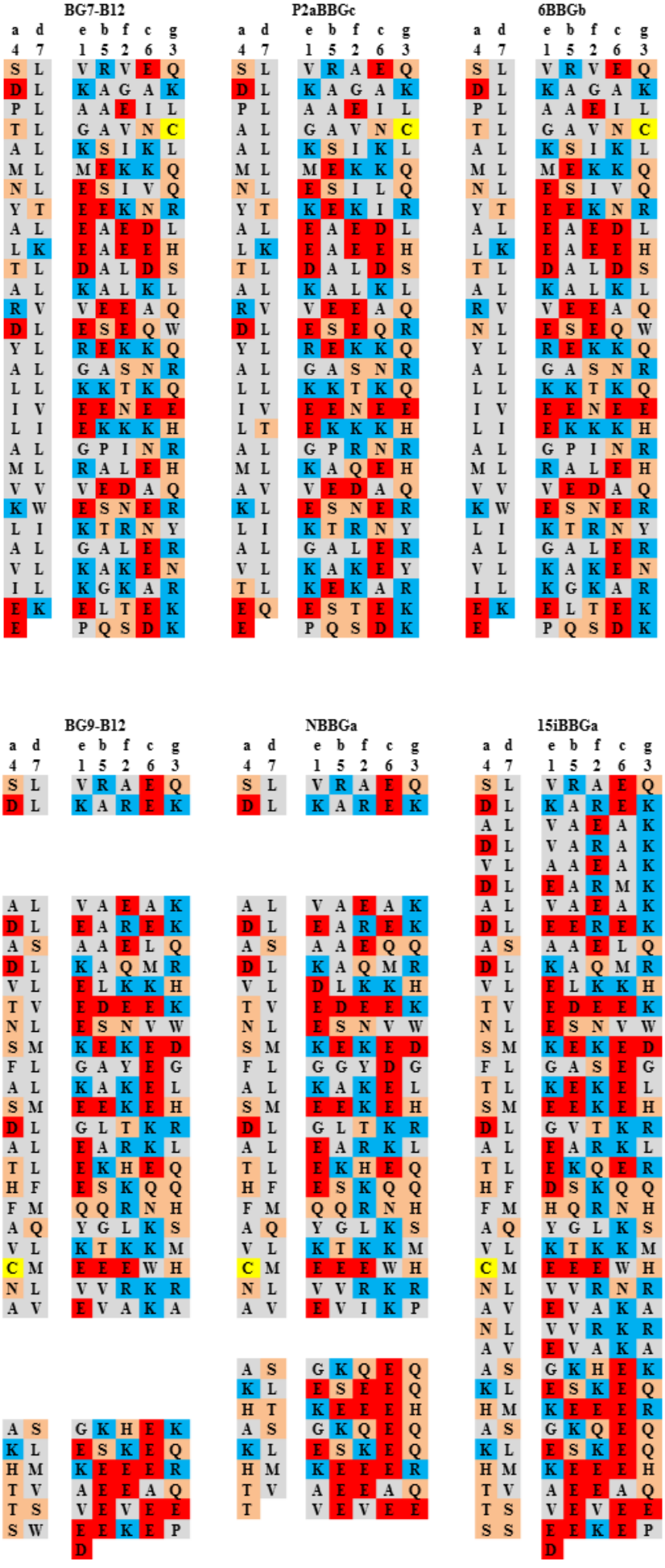


Figure legend is shown on the page following the figure

Figure 5.11 Coiled-coil representations of the cytoplasmic tails of all the dominantly expressed transcripts for each BG genes found in B cells from four chicken lines and BG7 and BG9 from B12 haplotype. The transmembrane region would be at the top of the page, so the C-terminus of the BG protein is at the bottom of the page. The positions of the seven codons in the 21 nucleotide repeat are indicated with numbers at the top, and the position of the seven amino acid positions of the 'true heptad repeat' are indicated with letters. Colours of squares surrounding the amino acids (single letter code) indicate features of the amino acids (red, acidic; blue, basic; orange, polar; grey, hydrophobic except for yellow, cysteine), with the understanding that these features do not correspond to full descriptions of the properties of the amino acids.

5.3.10 No obvious pattern was observed for the amino acid variation on cytoplasmic tails of these dominantly expressed BG genes

There was no obvious pattern observed for the amino acid variation on cytoplasmic tails of these dominantly expressed BG genes within the same BG clades. As shown in figure 5.12, all the variation coming from 6BBGb and P2aBBGc is annotated on the helical wheel structure of BG7; and all the variation coming from NBBGa and BG9 is annotated on the helical wheel structure of 15iBBGa. 15iBBGa has two extra insertions of a few heptad repeats compared to the other two, it is sensible to view all the sequence information on 15iBBGa.

Looking at the BG7 structure, there are 17 variable positions in total, with 2-3 found in every α-helical wheel position. All the variation is di-allelic; about two thirds of the variation is arguably conservative changes (T/A, I/T, V/T, W/L, R/K, V/A, L/Q, V/L, N/I, N/Y), while the rest is not (D/N, K/Q, E/K, G/E, I/R, and W/R).

Two kinds of variation exist among BG9, NBBGa, and 15iBBGa: the length of the cytoplasmic tails and the amino acid changes in certain positions. First of all, 15iBBGa has the longest tail, and the two insertions compared to the other two BG genes are labeled with light green blocks on 15iBBGa structure. The first insertion is a four heptad repeats after the second heptad repeat. The second insertion is a five heptad repeats if compared to BG9, or a two heptad repeats if compared to NBBGa, and it starts from the third heptad repeat after cysteine. Second, there are 20 amino acid changes among the three dominantly expressed BG genes. All the variation is scattered along the sequence but with only two changes in *a* and *d* positions. The variation is all di-allelic, most of which is arguably conservative changes (T/A, S/W, D/E, H/Q, A/G, S/Y, A/I, L/Q, A/P) with only a few that are not (E/A, N/K, R/Q, K/Q, R/H).

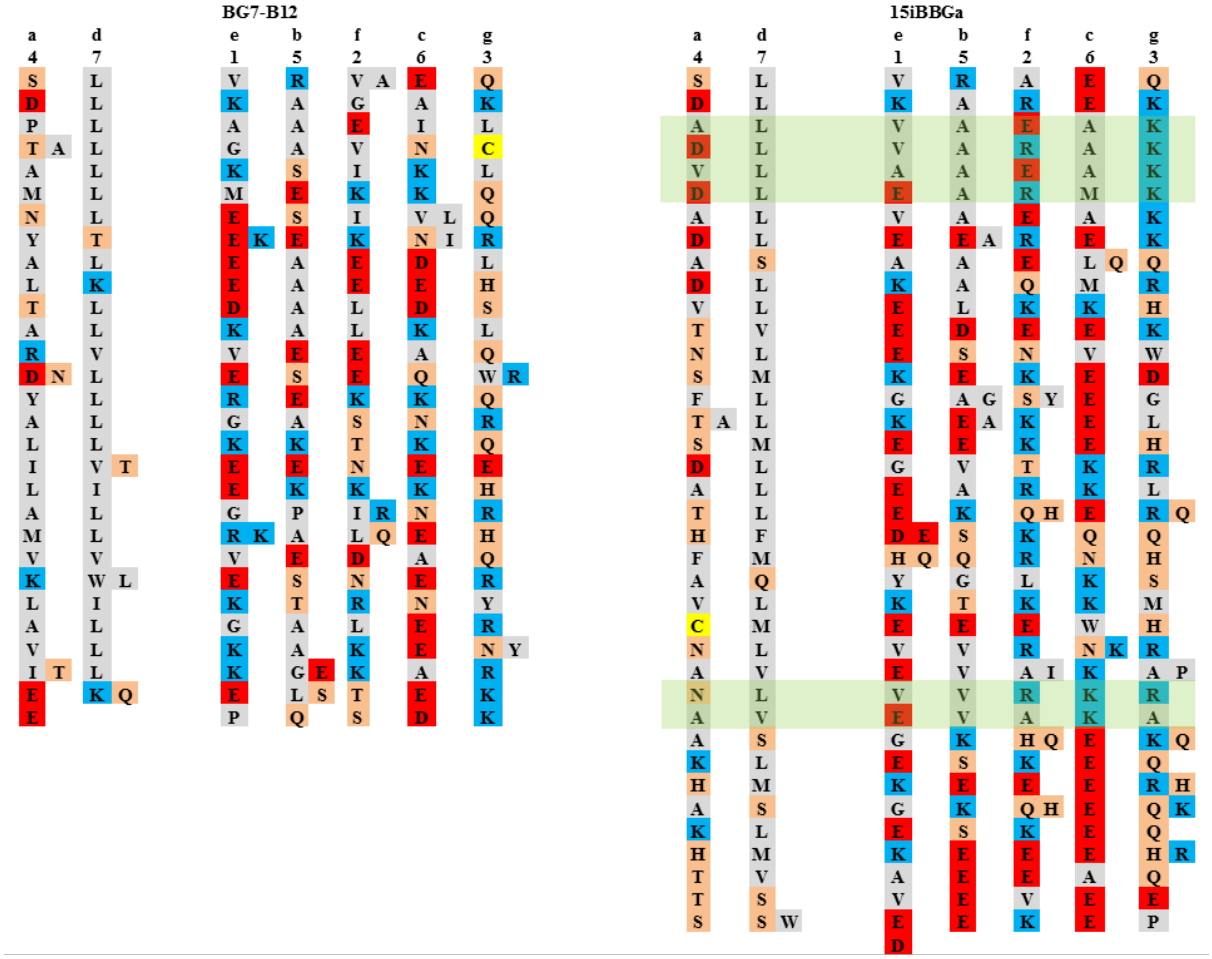
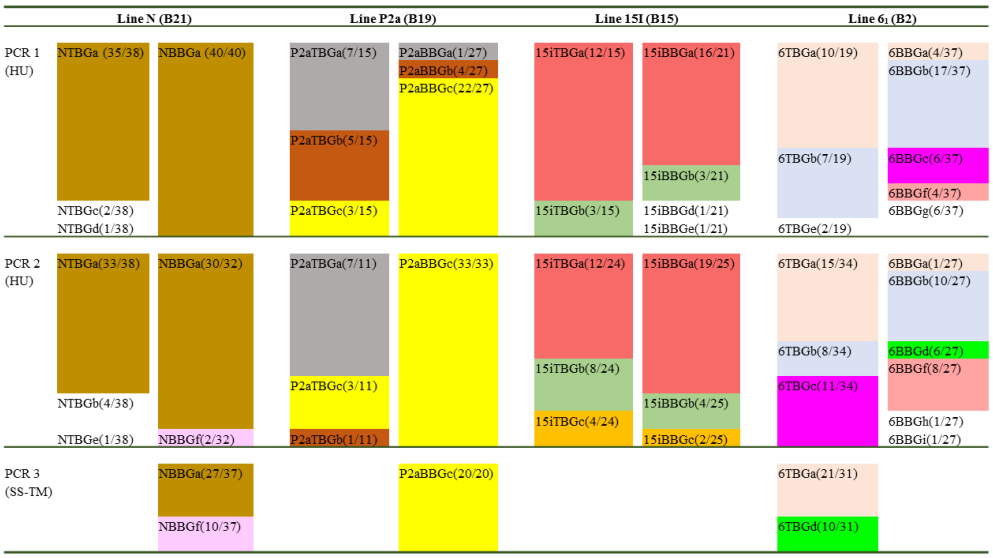


Figure 5.12 Coiled-coil representations of the cytoplasmic tails of the two representative BG genes of dominant expressed BG genes in B cells from BG8-9-12 clade and BG7-11 clade with variation annotated. The cytoplasmic tails of all the dominantly expressed BG genes found in the five chicken lines are BG7 and BG9 from line CB (B12), NBBGa from line N (B21), P2aBBGc from line P2a (B19), 15iBBGa from line 15I (B15) and 6BBGb from line 6_1 (B2). The cytoplasmic tail of BG7, P2aBBGc and 6BBGb belong to BG7-11 clade with the presentation made on BG7 sequence along with variation annotated on the left. BG9, NBBGa and 15iBBGa belong to BG8-9-12 clade with the presentation made on 15iBBGa sequence along with variation annotated and the insertions/deletions of heptad repeats squared by green labels on the right. The transmembrane region would be at the top of the page, so the C-terminus of the BG protein is at the bottom of the page. The positions of the seven codons in the 21 nucleotide repeat are indicated with numbers at the top, and the position of the seven amino acid positions of the 'true heptad repeat' are indicated with letters. Colours of squares surrounding the amino acids (single letter code) indicate features of the amino acids (red, acidic; blue, basic; orange, polar; grey, hydrophobic except for yellow, cysteine), with the understanding that these features do not correspond to full descriptions of the properties of the amino acids.

5.3.11 There are 23 different BG genes found in T and B cells from four chicken lines with nearly half expressed in both cell types

In order to understand the different and common features of the BG genes found in B cells and T cells, all 18 BG genes found in B cells, and all 16 BG genes previously found in T cells from four chicken lines were compared. In total, there are 23 different BG genes found in T and B cells from the four chicken lines, with 11 of these BG genes found in both cell types, while 12 found only in one cell type (Figure 5.13). It is worth noting that, in the figure 5.13 under each chicken line, the genes with the same small letters represent the same BG genes, and the difference in their names only tell from which cell type these BG genes were found. For example, NTBGa and NBBGa are the same BGa gene found in line N which have the exact same exon 2 sequence, and the former name indicates that it is from T cells, while the latter name from B cells.

The sequences with possible PCR artifacts were corrected during the comparison of all the cDNA sequences from both B and T cells. Previously, when analyzing the same BG genes from B cells or T cells separately, as there were not enough sequences to identify PCR artifacts, those cDNA sequences with various nucleotide variation all through the whole sequence except in exon 2 region (as exon 2 sequence is used to define different BG genes) were kept as independent transcripts under such genes. By comparing all the cDNA sequences of the same BG gene from T and B cells together, those possible PCR artifacts were picked and corrected accordingly. After correction, the transcripts found within the same cell type that have the same sequence were noted and only kept one name. To better understand this procedure, line N was used as an example in the following.



Line CB (B12)		Line N (B21)		Line P2a (B19)		Line 15I (B15)		Line 6₁ (B2)	
T cells	B cells	T cells	B cells	T cells	B cells	T cells	B cells	T cells	B cells
	BG7(16/40)	NTBGa(68/77)	NBBGa(97/109)	P2aTBGa(14/26)	P2aBBGa(1/80)	15iTBGa(24/39)	15iBBGa(35/46)	6TBGa(48/84)	6BBGa(5/63)
	BG8(2/40)	NTBGb(4/77)		P2aTBGb(6/26)	P2aBBGb(4/80)	15iTBGb(11/39)	15iBBGb(7/46)	6TBGb(15/84)	6BBGb(27/63)
BG9(25/28)	BG9(18/40)	NTBGc(2/77)		P2aTBGc(6/26)	P2aBBGc(75/80)	15iTBGc(4/39)	15iBBGc(4/46)	6TBGc(11/84)	6BBGc(6/63)
BG12(3/28)	BG12(1/40)	NTBGd(1/77)					15iBBGd(1/46)	6TBGd(10/84)	6BBGd(5/63)
	BG13(3/40)	NTBGe(1/77)					15iBBGe(1/46)	6TBGe(2/84)	
			NBBGf(12/109)						6BBGf(12/63)
									6BBGg(6/63)
									6BBGh(1/63)
									6BBGi(1/63)

Figure 5.13 Overall results for the number of BG genes amplified from T and B cells of four chicken lines. Top panel, independent amplifications from the four chicken lines; HU, haemopoietic forward and 'universal' reverse primers to give nearly full-length sequences; SS-TM, signal sequence forward and transmembrane reverse primers to give SS, extracellular Ig-V domain and TM regions. Different colours indicate different exon 2 sequences, except those sequences that are only found in one PCR. Names follow the convention: abbreviated line name, 'B' for B cells, 'T' for T cells, 'BG', and a letter representing the exon 2 sequence; numbers in parentheses indicate the number of clones found for a particular exon 2 sequence out of the total number for the particular PCR reaction. The same genes were labeled with the same colour. Bottom panel, the total results for four chicken lines from this project and for the CB line (B12) from Salomonsen *et al*., 2014.

As shown in appendix K, all BG cDNA sequences from line N (B21) were compared, and the sequence alignments between BG cDNA sequences from T cells and B cells helped to identify the nucleotide mutations resulting from PCR errors. Such obvious PCR errors labeled with blue in appendix K were corrected accordingly, and the modified transcripts were aligned and compared again for varification. As summarized in appendix L, the transcripts with the same sequence were noted with the same numbers in the 'Note' column. For example, NBBGa-1(x24), NBBGa-6(x2), NBBGa-13(x1), NTBGa-1(x35), NTBGa-2(x8), NTBGa-7(x1) and NTBGa-8(x1) would all have the same sequence, thus were noted with the same number '1'. Then these same transcripts were re-organized to generate the new names according to the following principles.

First of all, the old names of all the transcripts found in same cell type with the same sequences after correction were replaced with a new name to represent this sequence. The new name consists of three parts. First, the gene name followed by a dash and a number, in which the number is the same as the smallest number in the old names being replaced. Second, the total number of all colonies. Third, a number representing from how many PCR such sequence was found. For instance, as the example above, in B cells, NBBGa-1(x24), NBBGa-6(x2) and NBBGa-13(x1) have the same sequence; NBBGa-1(x24) has '1', the smallest number compared to '6' [NBBGa-6(x2)] and '13' [NBBGa-13(x1)], therefore, NBBGa-1 is the first part of the new name representing all three old names; the colony numbers of NBBGa-1(x24), NBBGa-6(x2) and NBBGa-13(x1), which are twenty four, two and one, respectively, are added into the new name; also, the sequence were found in two PCRs, thus the final name for all the three above is NBBGa-1(27, 2).

Secondly, if such sequences exist in both B cells and T cells, their final names included 'T' or 'B' to illustrate the tissue distribution, as well as to demonstrate that such sequences were found at least in two independent PCRs. For instance, as the same example above, NBBGa-1(27, 2) found in B cells has the same sequence as NTBGa-1(45, 2) found in T cells, therefore the final name for NBBGa-1 and NTBGa-1 were NBBGa-1(27, 2, T) and NTBGa-1 (45, 2, B), respectively.

This page intentionally left blank

5.3.12 Alternative splicing of BG transcripts differ in T and B cells

All the BG cDNA sequences from each chicken line were aligned against the genomic sequence of BG8 (line CB, B12 haplotype) to view the differences of transcripts in B cells and T cells. Alternative splicing is widely observed in the cytoplasmic tail of BG cDNA sequences from both B cells and T cells, primarily due to the intron retention, which leads to truncated cytoplasmic tails. However, the most striking finding is that only in B cells, some BG cDNA sequences are found with intron retained between exon 2 (encoding Ig-V domain) and exon 3 (encoding transmembrane region), resulting in an early stop codon introduced after the Ig-V domain, which in theory could be translated to a soluble protein.

5.3.12.1 Line N has the same BG gene dominantly expressed in B and T cells with lots alternative splicing isoforms but none encoding soluble BG protein

All the alternative splicing transcripts of BG gene found in line N (B21) from both T and B cells were compared and summarized in figure 5.14, with the details as follows.

First of all, the dominantly expressed BG gene in T and B cells is the same NTBBGa, with many different alternatively spliced transcripts (Figure 5.14). The dominantly expressed transcripts, NBBGa-1(27, 2, T) from B cells and NTBGa-1(45, 2, B) from T cells were the same sequence and were considered as the conceptual cDNA sequence of NTBBGa gene. There was no intron observed compared to other transcripts while all the other transcripts were the outcomes of alternative splicing.

Two factors together contribute to the alternative splicing: intron retention and the usage of alternative splice acceptor sites. Intron retention was only observed in the cytoplasmic tail regions, resulting in truncated cytoplasmic tails; and no transcript encoding soluble BG protein was observed in line N, unlike the other three chicken lines. The different usage of alternative splice acceptor sites was only observed in NBBGa-3 and NTBGa-4 (if the most dominantly expressed cDNA sequence NBBGa-1 and NTBGa-1 is treated as the standard transcript), and it happens at the penultimate exon of the cytoplasmic tail of NBBGa-3 and NTBGa-4, which although leads to one amino acid deletion, does not change the coding for the following amino acids.

Second, there were five BG genes only found in one of the two cell types. Four BG genes,

NTBGb, NTBGc, NTBGd, and NTBGe were found in one PCR with a few clones; while one BG gene, NBBGf, found in two PCRs but only in B cells. Not many alternative splicing isoforms were observed for these five BG genes, which might be due to the fact that few clones were detected.

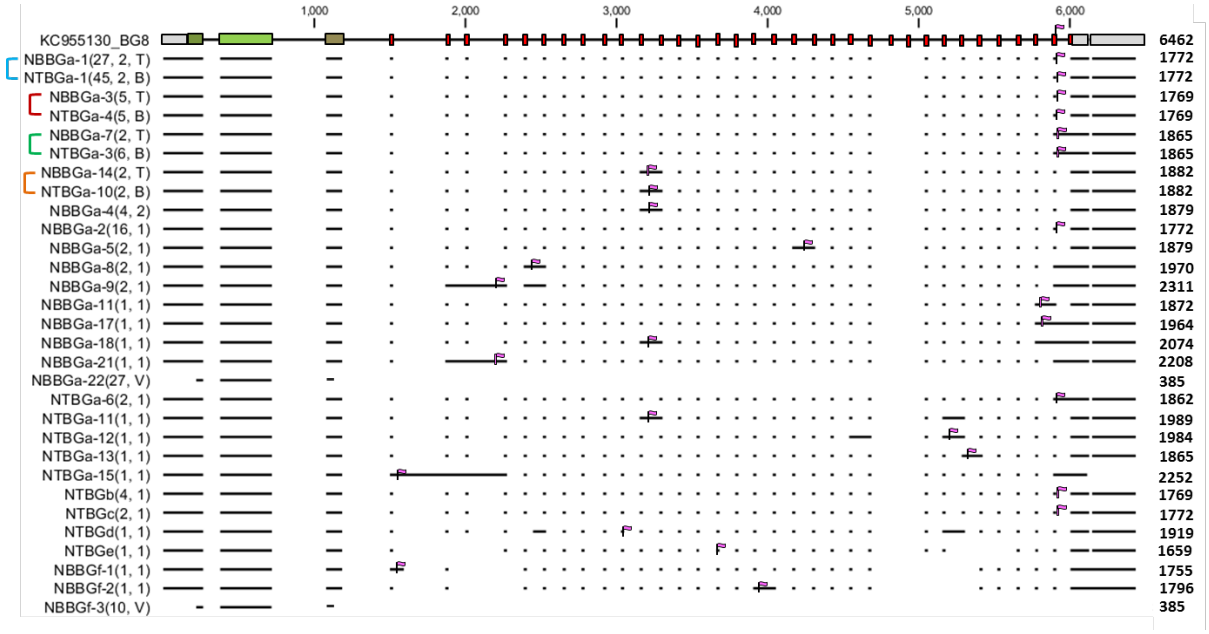


Figure 5.14 Representation of the intron-exon structure of the BG genes inferred from all the alternatively spliced transcripts found both in B and T cells from line N (B21). The top line represent the genomic sequence of BG8 from line CB (B12) with boxes indicating exons: grey for 5' UTR and 3' UTR, dark green for signal sequence, light green for Ig-V domain, brown for transmembrane, red for cytoplasmic tail. The actual mRNA transcripts are indicated as horizontal lines with stop codons labeled as vertical purple flags. Names of the transcripts follow the convention: abbreviated line name, 'B' for B cells, 'T' for T cells, 'BG' and the letter representing a particular exon 2 sequence, a dash and then a number representing the alternative splicing variant. Numbers in parentheses indicate the number of clones found for a particular full sequence, followed by the number of independent PCRs and/or a letter in which the sequence was identified (1, found in one PCR; 2, found in 2 PCRs; T, found in one PCR in the B cell project and one using T cell cDNA in previous T cell project; B, found in one PCR in the T cell project and one using B cell cDNA in the B cell project). The cDNA transcripts from B cells and T cells that have exact the same sequence are linked with a coloured bracket on the left side of the transcripts' names.

5.3.12.2 Line P2a has different dominantly-expressed BG genes in B and T cells, with transcripts encoding soluble BG protein only observed in B cells

There were only three BG genes found in line P2a (B19), and they were all detected in both cell types. However, unlike line N above, the dominantly expressed BG gene in B cells is different from that in T cells. Also, the transcripts encoding soluble BG protein were only observed in B cells. All cDNA sequences of BG genes from line P2a are discussed individually as follows (Figure 5.15).

P2aBGa, the dominantly expressed BG gene found in T cells but rarely found in B cells, did not show many different alternative splicing transcripts. Three different transcripts were found in T cells and a fourth transcript was found in B cells. By comparison to the genomic sequence of BG8 from line CB (B12), the dominantly expressed transcript, P2aTBGa-1(10, 2) has the complete full-length cDNA sequence without any intron. The other two transcripts in T cells both yield BG proteins with almost complete cytoplasmic tails, with P2aTBGa-3(1, 1) having one amino acid deletion due to the usage of alternative splice acceptor site, and P2aTBGa-2(3, 1) having a deletion of the last whole exon of cytoplasmic tail. The clone found in B cells, P2aBBGa(1, 1) has intron retention in the cytoplasmic tail region, which would lead to truncated cytoplasmic tail.

P2aBGb, like P2aBGa above, does not have alternative spliced isoform. The only transcript found in B cells has the exact same sequence as the dominantly expressed transcript found in T cells, which shows a full-length BG cDNA sequence without any intron. The other two transcripts found in T cells have introns retained at different locations, leading to different truncated cytoplasmic tails.

P2aBGc has many different alternative splicing isoforms, compared to the other two genes above. Strikingly, half of the clones (27 out of 54 in total) found in B cells have the intron retained between exon 2 (encoding Ig-V domain) and exon 3 (encoding transmembrane region). This introduces an early stop codon after the Ig-V domain and results in potential production of soluble protein. However, such cDNA sequences encoding soluble BG proteins were not seen in T cells. In T cells all the alternative splicing isoforms were seen in the cytoplasmic tail region, mostly due to intron retention leading to truncated cytoplasmic tails. Also, unlike the other two BG genes above, it seems that only one clone of P2aBGc from B cells, of which the cDNA sequence is named P2aBBGc-11(1, T), shows the full-length cDNA

sequence without any intron retained; while all the other cDNA sequences of P2aBGc found in B cells are alternative splicing isoforms.

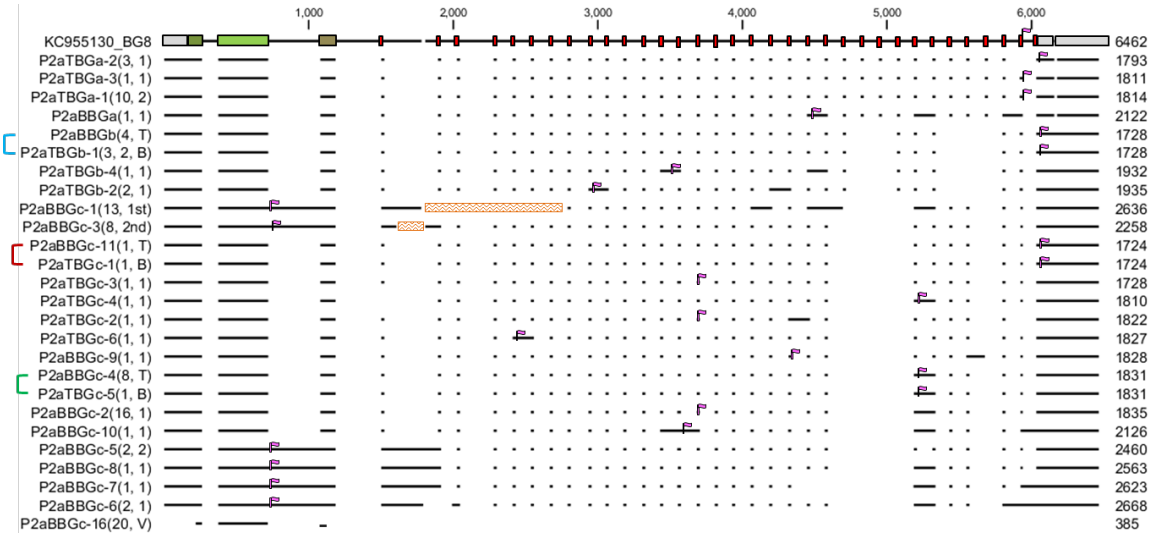


Figure 5.15 Representation of the intron-exon structure of the BG genes inferred from all the alternatively spliced transcripts found both in B and T cells from line P2a (B19). Same as legend to figure 5.14. Note: 1st, found in the first PCR; 2nd, found in the second PCR. The orange blocks in P2aBBGc-1(13, 1st) and P2aBBGc-3(8, 2nd) represent the regions that were not fully sequenced thus the sequence information was missed.

5.3.12.3 Line 15I has the same BG gene with the same transcript dominantly expressed in B and T cells, with transcripts encoding soluble BG protein only observed in B cells

For line 15I (B15), the 15iTBBGa was the dominantly expressed BG gene found in both cell types, with far more transcripts found for 15iTBBGa gene than any of the other four BG genes (Figure 5.16). There are three major points. One, unlike other chicken lines, most of the transcripts found in B cells were also found in T cells with long cytoplasmic tails. Two, the unique transcripts found in T cells encode shorter cytoplasmic tails compared to those in B cells. Three, only one transcript with three clones potentially encoding soluble BG protein was detected in B cells.

There were sub-dominantly expressed BG genes, 15iTBBGb and 15iTBBGc, which were both expressed in B and T cells. Two transcripts were found for 15iBBGb each with only three clones but both from two PCRs and they were found in T cells as well. These two transcripts encode long cytoplasmic tails, one with the full-length cytoplasmic tail (the same as the dominantly expressed transcript found in 15iTBGb), while the other one with the intron retained after the penultimate exon in the cytoplasmic tail (the same as the other transcript found in 15iTBGb), which added 14 amino acids into the tail and thus was even longer than the full-length cytoplasmic tail. There were three more transcripts found for 15iTBGb in T cells, all with a few clones, encoding long cytoplasmic tails except one transcript truncated cytoplasmic tail about half of the full-length. As for 15iTBBGc gene, only one transcript was found in B cells in one PCR with full-length cytoplasmic tail, and another transcript was found in T cells also from one PCR with truncated cytoplasmic tail which was about half of the full-length.

Two BG genes, 15iBBGd and 15iBBGe, were only found in B cells, each with only one clone showing transcripts encoding a full-length cytoplasmic tail and a soluble protein, respectively.

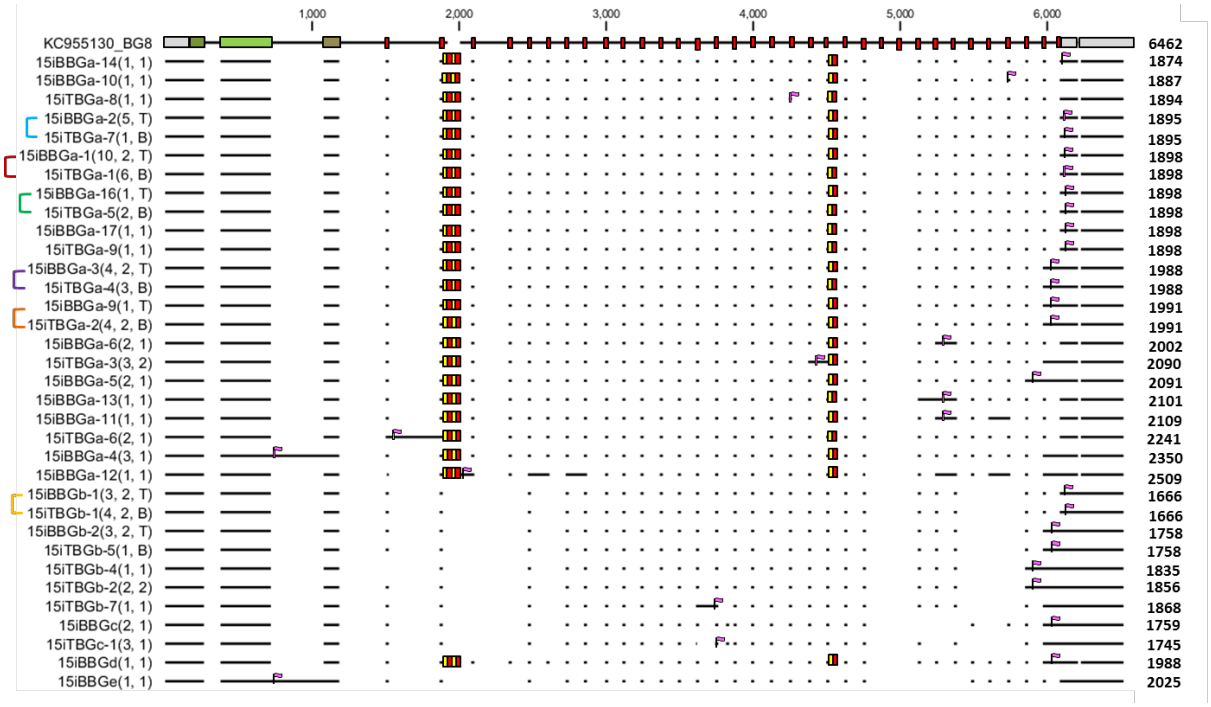


Figure 5.16 Representation of the intron-exon structure of the BG genes inferred from all the alternatively spliced transcripts found both in B and T cells from line 15I (B15). Same as legend to figure 5.14. Note: The alternating red and yellow boxes indicate the additional heptad repeats found for this gene, compared to BG8 (B12).

5.3.12.4 Line 6_1 has different BG genes strongly expressed in B and T cells, with short cytoplasmic tails in T cells and soluble proteins in B cells

There were nine different BG genes found in line 6_1 (B2) with five from T cells, eight from B cells, and four of them detected from both cell types; the dominantly expressed BG gene in T cells (6TBGa) is different from that in B cells (6BBGb) with striking alternative splicing isoforms. Like other chicken lines, the transcripts encoding soluble BG proteins are also only observed in B cells (Figure 5.17).

6TBGa was the dominantly expressed BG gene found in T cells, and all the transcripts have the intron retained after the first exon of the cytoplasmic tail region, resulting in very short truncated cytoplasmic tails with only 20 amino acids. The same gene was also found in B cells, but all the transcripts found encoded soluble proteins. 6TBBGa gene was the only gene among the 23 genes from which many transcripts were found, but none of them encoded normal full-length cytoplasmic tails. 6TBBGa has the same conceptual cDNA sequence (that is, exons without introns) as BG13 in B12 haplotype; however, it is not clear what the real transcripts look like for BG13 expressed in B12 cells.

6BBGb was the dominantly expressed BG gene in B cells. The most frequently found transcripts for 6TBBGb from T and B cells were the same, which encoded full-length cytoplasmic tails. Three more transcripts from 6TBBGb with a few clones, each encoded different truncated cytoplasmic tails.

The major transcripts for 6TBBGc in T and B cells were the same sequence encoding full-length cytoplasmic tails. The other gene detected in both B and T cells, 6TBBGd, was found as a single transcript in B cells which encoded a truncated cytoplasmic tail. In T cells, 6TBBGd was only analyzed by SS-TM primers, therefore its sequence only covers partial signal sequence, the whole Ig-V domain and partial transmembrane region.

The other five BG genes only found in one of the two cell types were only detected in one PCR. Since there were only few clones picked, the number of transcripts might be biased. However, with limited information, transcripts from two BG genes, 6BBGg and 6BBGi, both encoded a very short truncated cytoplasmic tails, due to the intron retained after the first exon of cytoplasmic tail, the same situation as 6TBGa gene in T cells described above.

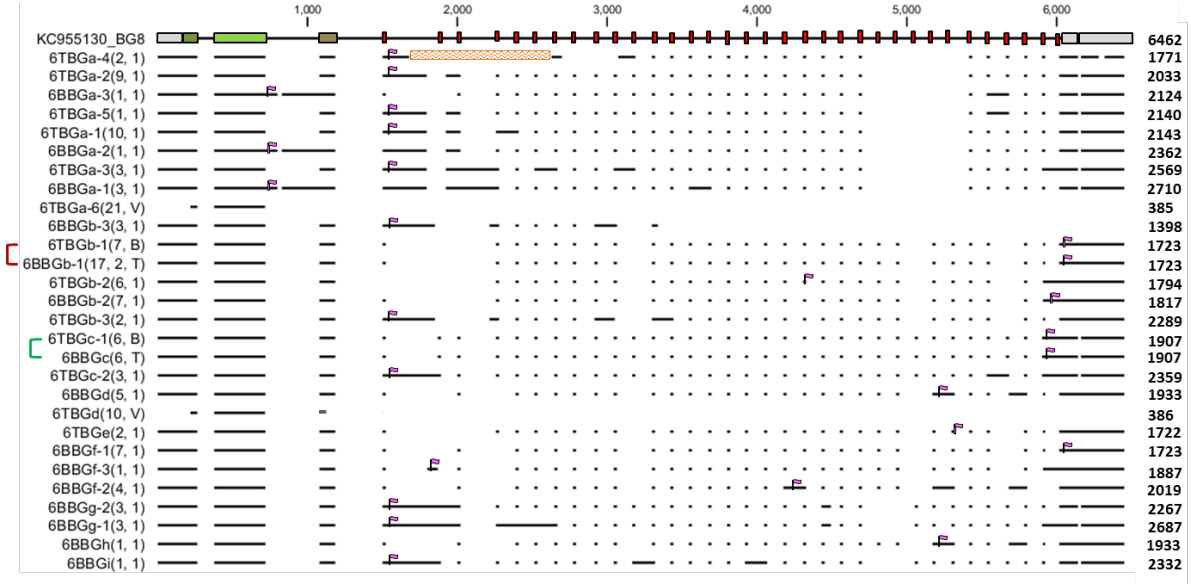


Figure 5.17 Representation of the intron-exon structure of the BG genes inferred from all the alternatively spliced transcripts found both in B and T cells from line 6_1 (B2). Same as legend to figure 5.14. Note: the orange block in 6TBGa-4(2, 1) represents the region that was not fully sequenced thus the sequence information was missed.

5.3.12.5 Summary

To summarize, all the cDNA sequences of BG genes found in T and B cells have alternative splicing isoforms, mostly due to intron retention. Some cDNA sequences found in B cells have the intron retained between exon 2 (encoding Ig-V domain) and exon 3 (encoding transmembrane region), resulting in potential production of soluble BG proteins. Most genes expressed transcripts encoding full-length cytoplasmic tails without intron retention, but some genes were only found with transcripts encoding very short truncated cytoplasmic tails in either or both cell types, or soluble proteins in B cells. In some circumstances that some transcripts found with very few clones in T cells but with many clones in B cells (vice versa) might be caused by cell contamination during cell sorting.

5.3.13 BG genes are highly polymorphic but same BG genes have the same cDNA sequences from different chicken with the same B haplotypes

Thus far, we have compared all the BG genes found from T and B cells of four chicken lines, line N (B21), line P2a (B19), line 15I (B15), line 6_1 (B2), and examined in detail various alternative splicing transcripts of these BG genes. However, we haven't compared our sequences to the ones found by other groups or from previous work done in our lab. Therefore, in this section, as many BG genes as we could find from GenBank and ENSEMBL were collected and compared to our sequences, in order to understand whether BG genes are conserved between different chickens with the same MHC haplotype, and what the phylogenetic relationships of these genes with ours are.

5.3.13.1 Comparison between BG genes from BQ (Gaga5.0) and B12 provides evidence for swapping BG clusters between haplotypes during evolution

Fifteen BG genes were annotated from Gaga5.0. In total, there were 25 genomic locations in Gaga5.0 with high homology to BG genes by using the nucleotide sequence of exon 2 (encoding Ig-V domain) of BG8 from B12 haplotype to BLAT the Gaga5.0 genome sequence in ENSEMBL. As shown in appendix M, eighteen locations are located on chromosome 16, one on chromosome 2 and another six on unlocalized scaffolds, virtually all with high E value. However, checking the sequences in detail, seven locations including three from chromosome 16 and four from unlocalized scaffolds only had a short fragment showing homology to Ig-V domain of BG8. BG genes are known to be polymorphic, but the genomic structures are very conserved, with exon 2 composed of 342 bp encoding Ig-V domain. Therefore, only those locations with the whole 342 bp were considered for further evaluation on whether they had typical BG gene structures. The typical BG gene should have one exon encoding 5' UTR and 99 bp signal sequences, one exon encoding Ig-V domain (342 bp), one exon encoding transmembrane region (105 bp), many small 21 nucleotide repeat exons encoding cytoplasmic tail, and one exon encoding 3' UTR. Eventually, 15 locations were confirmed to be BG genes and three locations turned out to be non-coding transcripts. The genomic locations of the 15 BG genes of BQ haplotype on Gaga5.0 are indicated in figure 5.18. Two singleton BG genes, WGS0-BQ and WGS1-BQ, are annotated on chromosome 2 and BF/BL region of chromosome 16 respectively. All 13 BG genes are in the BG region on chromosome 16 upstream of the TRIM region.

Genomic location comparison of BG genes in BG regions between BQ (Gaga5.0) and B12 haplotypes together with phylogenetic study on Ig-V domains showed evidence of swapping BG clusters between haplotypes. Previously Dr. Chattaway compared BQ (also from Gaga5.0 annotation) and B12 haplotype. He found three identical BG genes and suggested the swapping of whole BG clusters between the two haplotypes, B12 cluster IV and BQ Cluster I both labeled green in figure 5.18 (Chattaway, 2013). Due to the lack of sequence data of those annotated BG genes done by Dr. Chattaway, it is unclear which three genes are identical; also it is hard to compare his work to mine. However, our comparison result supports the proposal of BQ and B12 swapping the whole clusters (BQ Cluster I and B12 Cluster VI) during evolution. As shown both in the genomic comparison (Figure 5.18) and the phylogenetic tree built on the Ig-V domain (exon 2) nucleotide sequence alignments (Figure 5.19), two pairs of BG genes are found identical between BQ and B12 haplotypes, with another two pairs having very similar sequences. The two identical BG pairs are BG12-WGS5 in the B12 Cluster VI - BQ Cluster I region, and BG0-WGS0 on chromosome 2. The two similar pairs are BG10-WGS3 in the B12 Cluster VI - BQ Cluster I region, and BG1-WGS1 in the BF/BL (MHC) region. However, according to the phylogenetic tree, WGS2 from BQ Cluster I is clustered with BG3-4-5 in a big clade, and BG13 from the B12 Cluster VI is grouped with WGS14, suggesting the two BG genes, WGS2 from BQ Cluster I and BG13 from B12 Cluster VI, might not be involved in the swapping event.

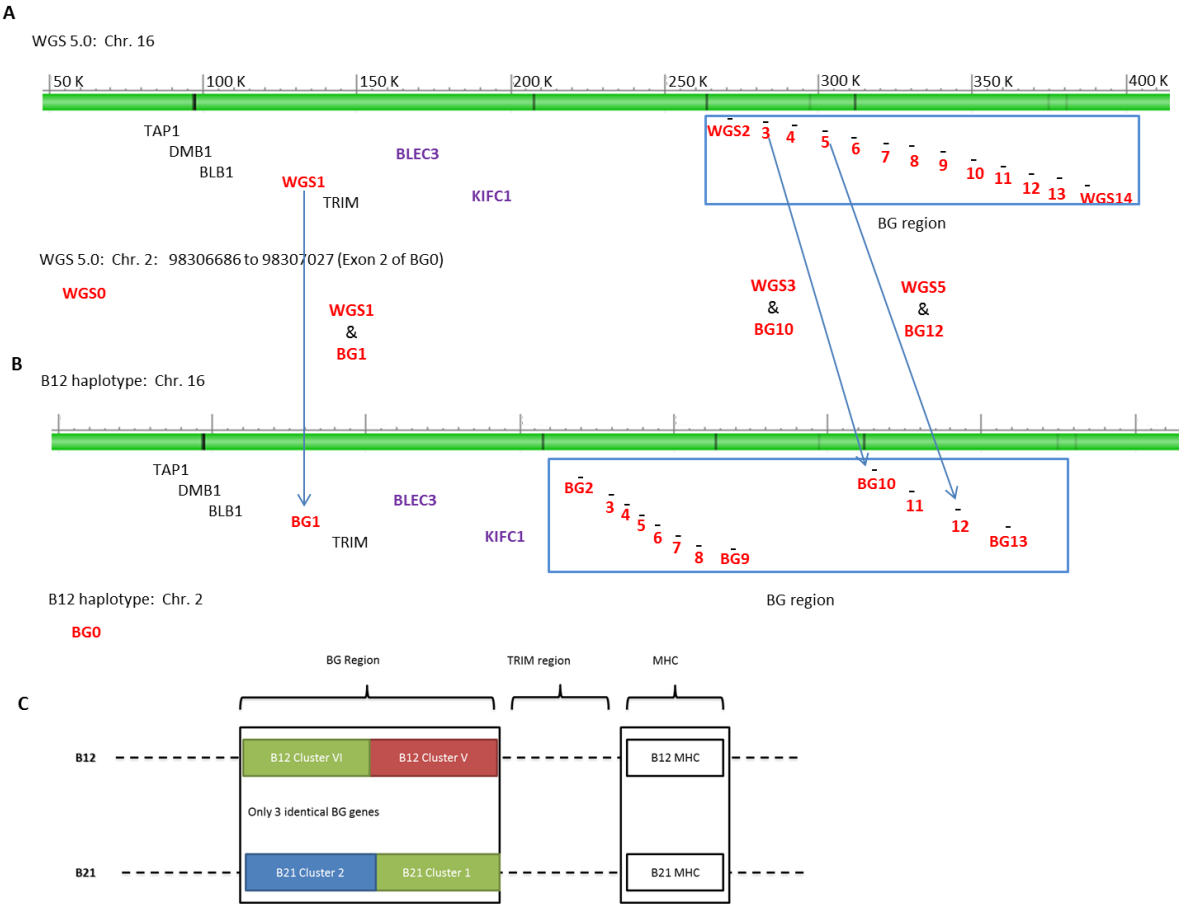


Figure 5.18 Comparison of genomic locations of 15 BG genes annotated on chicken whole genome sequence (Gaga5.0) of BQ haplotype and 14 BG genes on B12 haplotypes. A. Annotations of all 15 predicted BG genes on whole genome sequence of a female red jungle fowl Gaga5.0 with MHC BQ haplotype (line UCD 001). All the BG genes (from WGS0 to WGS14) are labeled in red and the BG region was captured in dark blue rectangular box. B. An indication of general genomic locations of all 14 BG genes (from BG0 to BG13) in B12 haplotype. There are two singleton BG genes on Both BQ and B12 haplotypes, one on chromosome 2 and the other one in the BF/BL region on chromosome 16. WGS1 has the identical exon 2 sequence as BG1, and so do WGS5 and BG12; while WGS3 and BG10 are clustered together in the phylogenetic tree built on exon 2 sequences of all WGS BG genes and B12 BG genes. C. Illustration of the BG cluster in green shared by B21 and B12 haplotype (diagram copied from Prof. Kaufman).

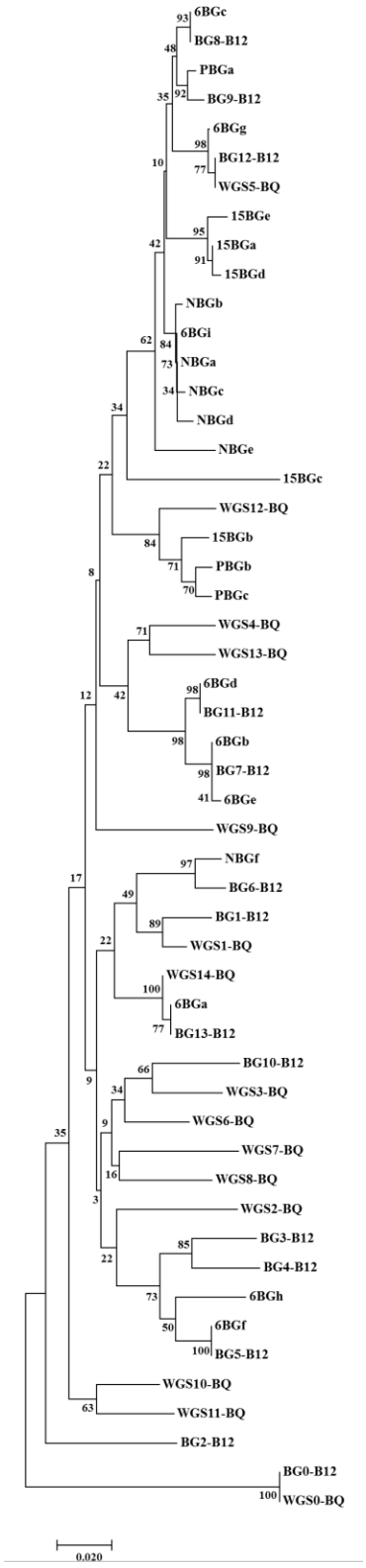


Figure 5.19 Phylogenetic tree built on nucleotide sequences of exon 2 of 15 BG genes annotated on chicken WGS (Gaga5.0) from BQ haplotype and all 14 BG genes from B12 haplotype.

5.3.13.2 Two different chickens from line 15I carry the same BG genes

The two BG genes found in chapter 3 by amplification using H2 primers from duodenum of line 15I chicken were the same transcripts of 15iTBBGa found in both B and T cells from another line 15I chicken here by HU primers, indicating the BG genes are the same in different chickens from the same chicken line and B haplotype. Nucleotide sequence alignment (Figure 5.20) shows that B15DLBG1 has exactly the same sequence as the most dominantly expressed transcripts in both B and T cells, 15iTBGa-1(6, B) and 15iBBGa-1(10, 2, T); while B15DLBG2 is virtually identical to a transcript of 15iTBBGa [15iTBGa-2(4, 2, B)] in sequence (with one nucleotide difference in the Ig-V domain, which is possibly due to PCR error). Previously in chapter 3, not many alternative splicing transcripts were observed, therefore, the large insertion in the cytoplasmic tail and the one nucleotide difference in Ig-V domain were considered to be from a different gene. However, after systematic examination of the cDNA sequences of BG genes, it is clear that the two BG cDNA sequences found from line 15I chicken are most likely to be the same as the one we found in another line 15I chicken.

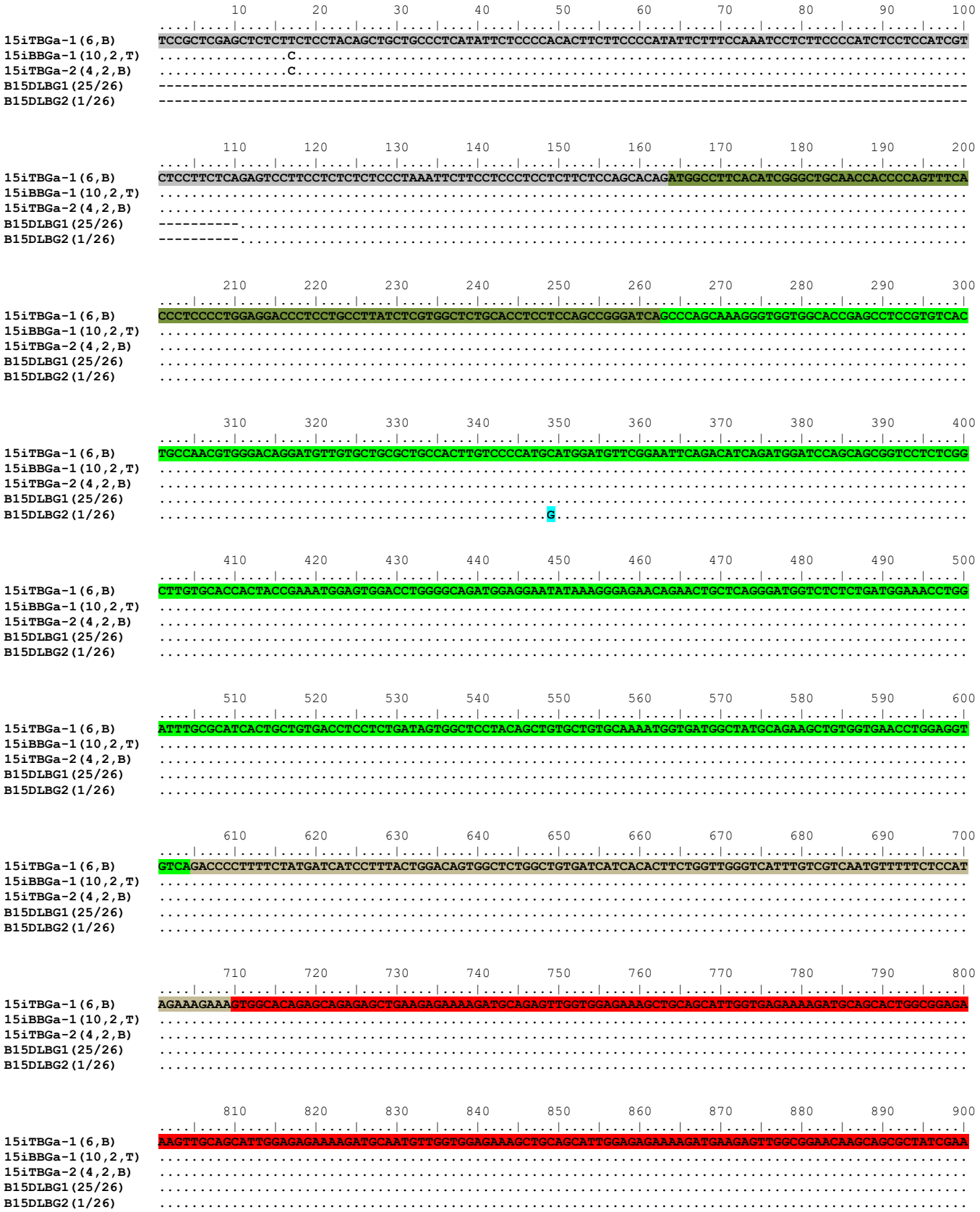


Figure legend is shown on the page following the figure

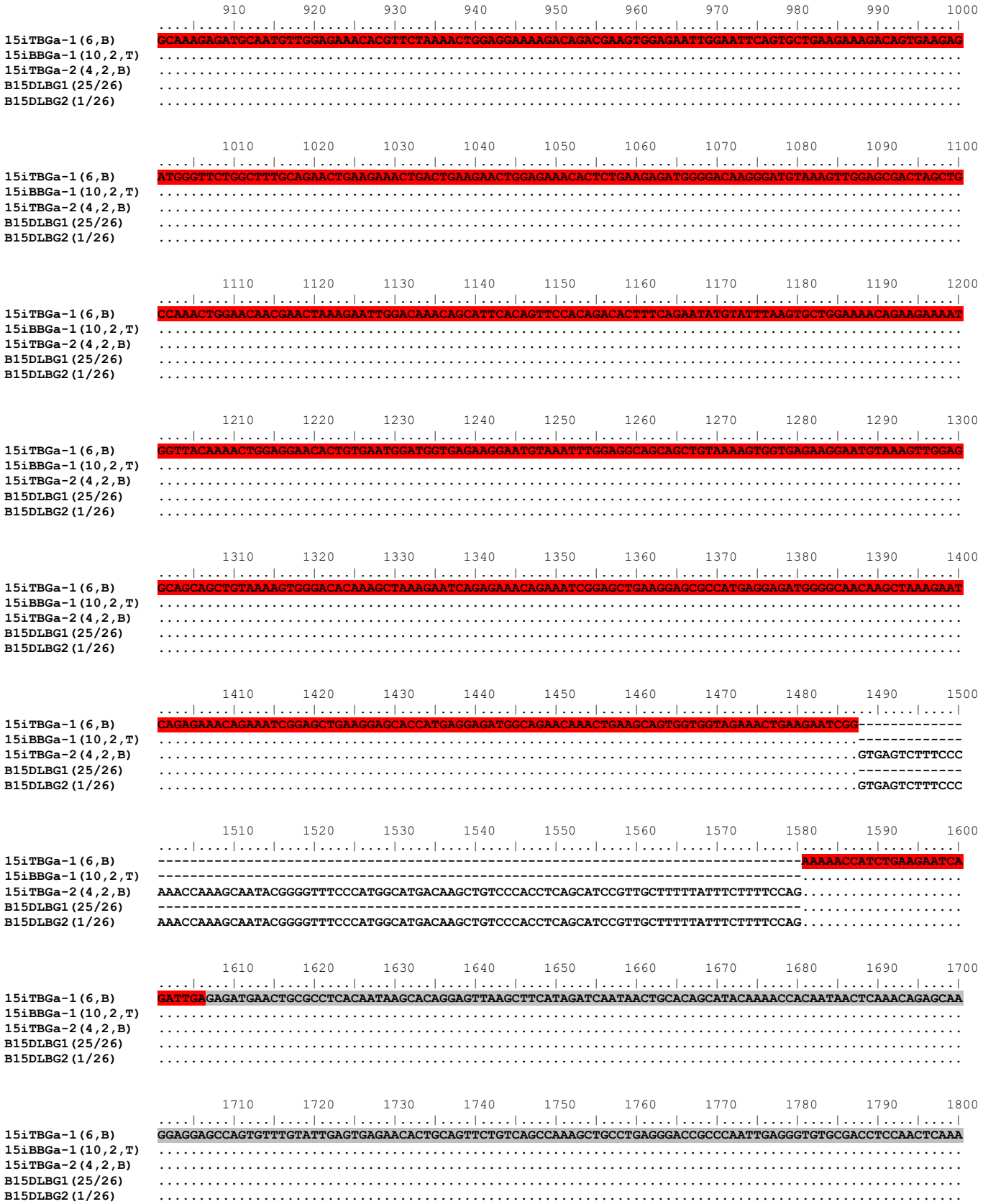


Figure legend is shown on the page following the figure

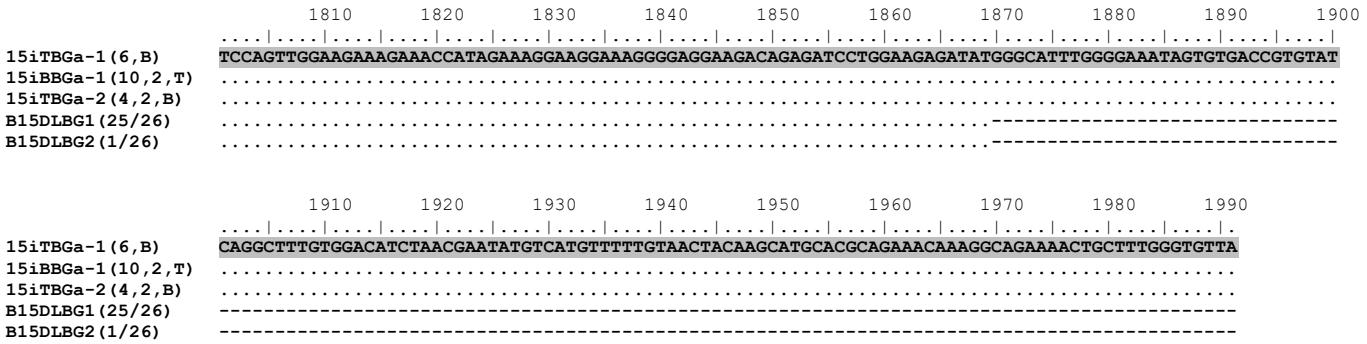


Figure 5.20 Nucleotide sequence alignment of BG transcripts found in two individual chickens but both line 15I (B15). The two BG genes found previously by H2 primers in duodenum sample of B15 haplotype chicken was actually the same gene as 15TBBGa found in T cells and B cells from line 15I (B15) using HU primers. Nucleotide sequence alignment shows that B15DLBG1 has the exactly same sequence with 15TBBGa, while B15DLBG1 should be 15TBBGa gene as well with alternative splicing in the cytoplasmic tail region and one nucleotide difference in the Ig-V domain which is likely to have been caused by PCR artifact. The colours labeled on top sequence indicate the regions of its cDNA structure: 5' UTR and 3' UTR are shown in grey, signal peptide in dark green, Ig-like V domain in bright green, transmembrane region in dark brown, and cytoplasmic tail in red.

5.3.13.3 Most of the BG cDNA sequences of line UCD330 (B21) downloaded from GenBank are the same as the BG cDNA sequences found in line N (B21)

In order to understand what the BG cDNA sequences in GenBank are and what the relationships between these sequences and the ones found in my project are, all the eight nearly full-length BG cDNA sequences and the one containing only partial BG cDNA sequence were aligned with the conceptual cDNA sequences of all 14 B12 BG genes and 23 BG genes found from four different haplotypes (B21, B19, B15 and B2) in this project. As summarized in table 5.1 previously, most of these cDNAs downloaded from GenBank were from line UCD330 as described by the authors, which should be B21 haplotype.

Therefore, we expected that some of the UCD330 sequences should match with the sequences here we found from line N, because they are both B21 haplotype. However, it would not be surprising if there were some sequences we didn't see in line N, because only the haemopoietic BG genes were explored in this project and unmatched sequences might be tissue BG genes.

Not all the BG cDNA sequences downloaded from GenBank were full length

As shown in appendix N, all BG cDNA sequences from GenBank have shorter 5' UTRs compared to our sequences, but have longer 3' UTRs except two sequences, U60-Unk and NM-B21. Only four B21 sequences from GenBank are long enough in 5' UTRs to be confirmed as haemopoietic BG sequences by phylogenetic tree (Figure 5.21) as well as the obvious deleted fragment in the 5' UTRs compared to tissue BGs in the nucleotide sequence alignments (Appendix N) (Salomonsen *et al.*, 2014). As for the extra nucleotides in the 3' UTR regions of these BG cDNA sequences downloaded from GenBank, they are about 50 bp longer than our sequences and contain multiple poly (A) signal sites (AATAAA). Upstream of the first AATAAA motif, there are only 22 bp extra nucleotides in these GenBack cDNA sequences compared to ours. Therefore, the HU primers did perform well and could amplify nearly full-length cDNA of BG gene. For the accuracy of further phylogenetic analysis, these extra 3' UTR nucleotides in the GenBank cDNA sequences were deleted in the alignments.

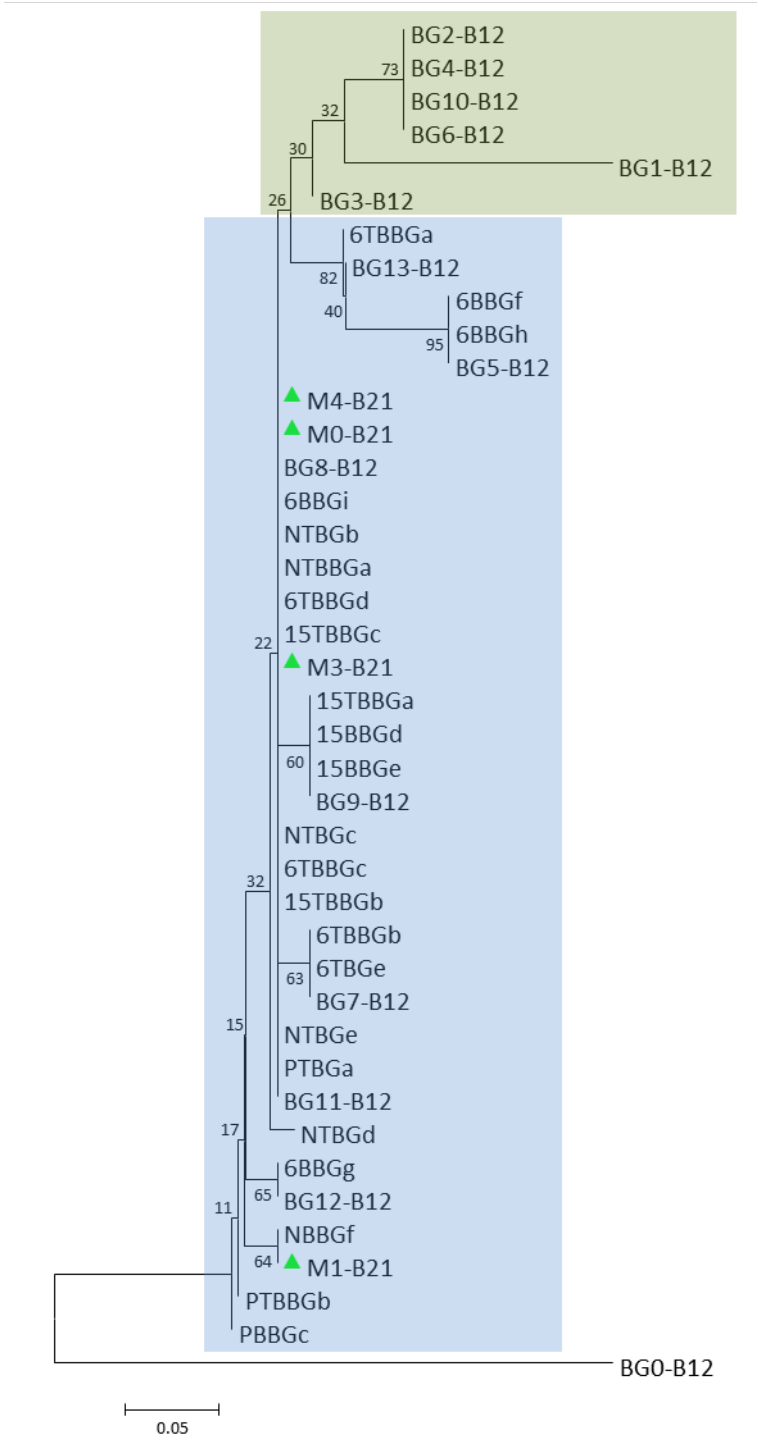


Figure 5.21 Phylogenetic tree built on 5' UTRs showed that four B21 cDNA sequences downloaded from GenBank belong to haemopoietic BGs. The sequences could be divided into two major groups, the group containing tissue BGs (green) and the group containing haemopoietic BGs (blue) according to Salomonsen *et al*., 2014. Other GenBank BG sequences were too short or lack 5' UTR information to be compared. The green triangle helps to distinguish the four B21 cDNA sequences from others.

Four GenBank B21 BG sequences were the same gene as NTBBGa gene in line N (B21)

Nucleotide sequence alignments clearly show that four BG cDNA sequences, M0-B21, M3-B21 and M4-B21 from Miller et al (Miller *et al.*, 1991), and U60-Unk from Goto et al (Goto *et al.*, 1988), were the same BG gene as NTBBGa. There are very few differences in the cytoplasmic tail and 3' UTR regions compared to the conceptual cDNA sequence of NTBBGa, except that U60-Unk arguably only contained a partial cDNA sequence (Appendix O). The background about these four BG cDNA sequences is described in Material and Methods in section 5.2.5.3. As shown in appendix O, the differences between these four BG cDNA sequences and the conceptual cDNA sequence of NTBBGa are due to either PCR errors or evolutionary nucleotide variation, as well as alternative splicing and small fragment deletion in the cytoplasmic tail.

Two GenBank B21 sequences were very close to NBBGf in line N (B21)

Two GenBank BG cDNA sequences, M1-B21 and NM-B21 both from a cDNA library made using line UCD330 (B21) erythroid cells (Miller *et al.*, 1991), were very similar to NBBGf found in our project with only a few differences. As shown in the nucleotide sequence alignments between the conceptual cDNA sequence of NBBGf and the two GenBank BG cDNA sequences (Apeendix P), M1-B21 and NM-B21 are virtually identical except there is no 3' UTR nor 5' UTR information for NM-B21. There is a six nucleotide repeat at the start of signal sequence in M1-B21, which might be a PCR artifact. Also, one nucleotide difference in the Ig-V domain, as well as the one nucleotide difference in the cytoplasmic tail, might both be due to PCR errors.

Another two GenBank sequences have type 1 cytoplasmic tails and were close to BG10 and the BG7-11clade of B12 haplotype respectively

A phylogenetic tree built from the regions covering the Ig-V domains to the end of 3' UTRs of all available cDNA sequences showed that the other two BG cDNA sequences, U98-Unk from Bikle et al (Bikle *et al*., 1996) and M2-B21 from Miller et al (Miller *et al.*, 1991), were clustered with BG10 and BG7-11 clade of B12 haplotype, respectively (Figure 5.22). Nucleotide sequence alignments between these two BG cDNA sequences with all the 14 BG genes from B12 haplotype clearly show that they both have type 1 cytoplasmic tails (Appendix Q).

U98-Unk was found in the intestine and was called the zipper protein by the author (Bikle *et al*., 1996). Nucleotide sequence alignments between U98-Unk and BG10-B12 (Figure 5.23) showed some variation in the Ig-V domains and 3' UTR regions (Salomonsen *et al.*, 2014), but they are quite conserved in the cytoplasmic tails where an 18 nucleotide fragment missing in U98-Unk might be due to alternative splicing.

M2-B21, unlike other B21 cDNA sequences investigated here, was not obviously similar to any BG genes found in line N or other chicken lines, but was clustered in the BG7-11 group of B12 haplotype (Figure 5.22). Also because M2-B21 lacked 5' UTR sequence information, there was no clue whether it was a haemopoietic BG gene or tissue BG gene.

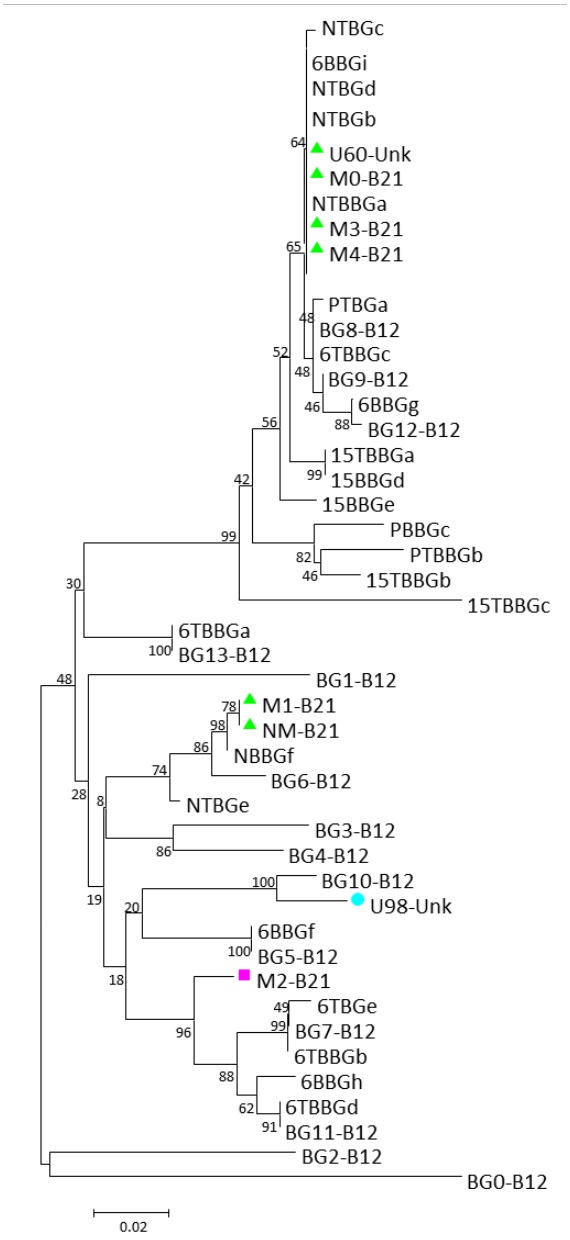


Figure 5.22 Phylogenetic relationships between eight BG genes from GenBank, all 23 BG genes found in this project and 14 BG genes from B12 haplotype. Phylogenetic tree is built on nucleotide sequences from Ig-V domains to 3' UTRs. The eight BG cDNA sequences from GenBank fall into different clusters, with four belonging to BG8-9-12 group, two closely related NBBGf and BG6, one grouped with BG10, and one going to BG7-11 group. The green triangle, blue round and pink square help to distinguish the BG cDNA sequences downloaded from GenBank from others.

Figure legend is shown on the page following the figure

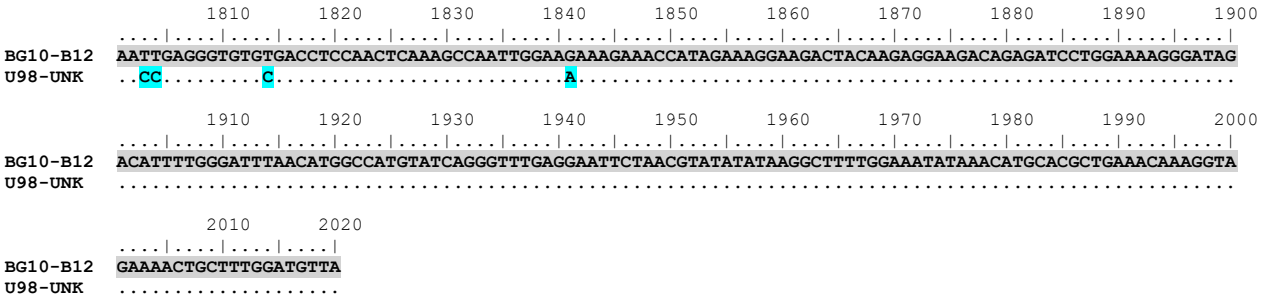


Figure 5.23 Nucleotide sequence alignment shows that the 'zipper' protein and BG10 from B12 are highly conserved in cytoplasmic tails. U98-Unk was called zipper protein or BG2 by its author. The differences between 'zipper' protein and BG10-B12 are labeled blue on 'zipper' protein sequence whose protein translation stop codon was labeled in purple. The cDNA structure is indicated with different colour labels on BG10-B12: 5' UTR and 3' UTR in grey, signal peptide in dark green, Ig-V domain in bright green, transmembrane region in dark brown and cytoplasmic tail in red. Same alignment has been done previously in Salomonsen *et al*., 2014.

5.3.13.4 Ig-V domains of BG genes from free-range broiler chickens showed high polymorphism but most of them within known BG clades

A phylogenetic tree built on the Ig-V domains of all the Comperos chickens and other BG cDNA sequences downloaded from GenBank, as well as the 23 BG genes from four chicken lines and 14 BG genes from B12 haplotype showed high polymorphism but most of the sequences falling into the known clades (Figure 5.24). It was discussed above that some BG cDNA sequences downloaded from GenBank are the same genes as those found in this project, and as they were from the same chicken B haplotype, it is suggested that BG genes from the same B haplotype are conserved in different chickens. Also, the 23 BG genes found from four chicken lines in this project are different to each other, but some genes from one chicken line (line 6_1, B2 haplotype) are the same as those in B12 haplotype. So far, it has been unclear what the BG genes from outbred chickens are like. The BG cDNA sequences with 'Cam' in the gene names from table 5.1 were from Cameros chicken, which were generated from 51 free-range broiler chickens by Iglesia et al (Iglesias *et al.*, 2003). The published sequences of these Cameros BG genes only contain exon 1 and exon 2, therefore, the Ig-V domains (exon 2) were aligned with all the other BG genes.

As shown in figure 5.24, none of the Cameros BG genes is identical to any other known BG gene, showing polymorphism of Ig-V domain. However, most (11 out of 16) of these Cameros genes fall into the BG clusters that have been well recognized in this project. Four sequences fall into BG7-11 clade, two in BG8-9-12 clade, two in Pbc15b clade, two with BG10, one with BG1, and the other five are not grouped with any particular clade or genes. None of the genes fall into BG3-4-5 clade which matches our findings about the 23 BG genes from four chicken lines, as both our samples and Cameros samples are from blood. However, we didn't find BG10 like gene in our sample since it is a tissue BG.

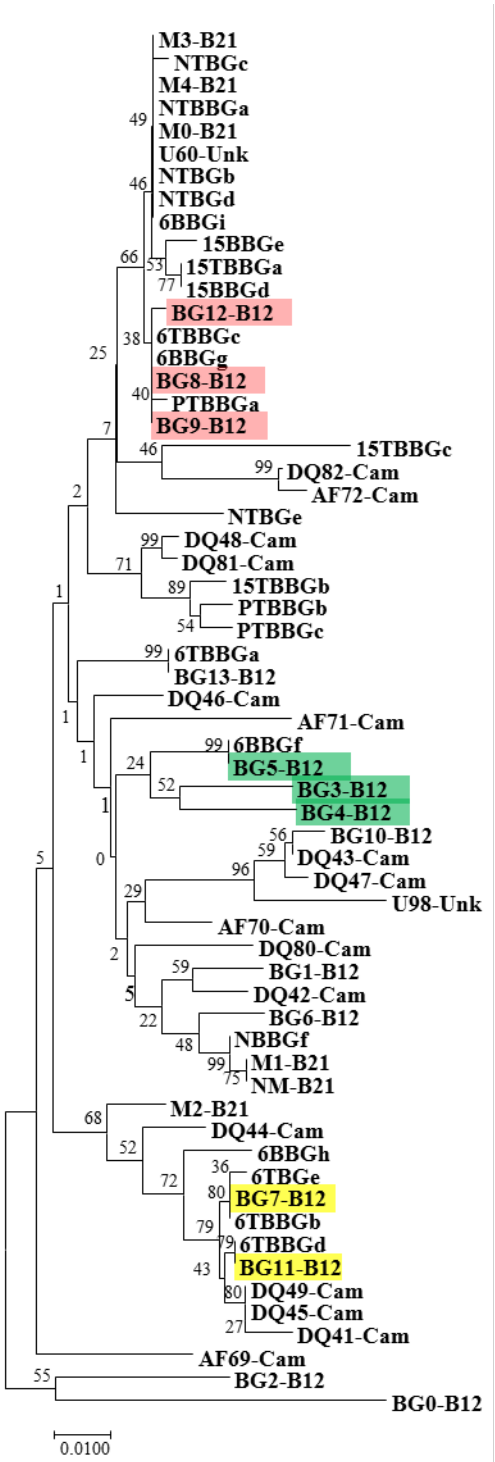


Figure 5.24 Phylogenetic relationships of all BG genes mentioned in this project. Phylogenetic tree is built on nucleotide sequences of the exon 2 (encoding Ig-V domain) for all the BG cDNA sequences discussed in this chapter, including all 23 BG genes found in T and B cells of four chicken lines, all 14 BG genes on B12 haplotype from Salomonsen et al 2014, all 15 BG genes annotated from chicken whole genome sequence (WGS), Gallus_gallus-5.0 (Gaga5.0), and all the BG cDNA sequences downloaded from GenBank listed in table 5.1. Same colours indicate that the genes are clustered in the same clade in the phylogenetic tree built on exon 2 region of B12 BG genes from Salomonsen et al., 2014.

5.4 Discussion

In this chapter, we examined the BG cDNA sequences in B cells from four chicken lines, line N (B21), line P2a (B19), line 15I (B15) and line 6_1 (B2), using the same methodology applied in previous T cell project. The results enrich our understanding about BG genes in the following aspects.

First of all, a total of 18 BG genes (based on exon 2 sequence) was found in B cells with 11 of the genes found previously in T cells from the same individual chickens. In line N, two BG genes were found in B cells versus five BG genes in T cells, and only one gene was found in both cell types. If we use 'B2:T5:BT1' to represent such results from line N, then the results for the other three chicken lines are: 'B3:T3:BT3' for line P2a, 'B5:T3:BT3' for line 15I, and 'B8:T5:BT4' for line 6_1. Together there are 23 BG genes found in these four chicken lines in both B and T cells, and none of them is identical to each other. However, six BG genes out of nine found in line 6_1 are the same genes as those found in B12 haplotype, which helps to support the previous observations that B2 and B12 haplotypes share some serological identity on erythrocytes (Simonsen *et al.*, 1982) and that 2-D gels of BG antigens made from B2 and B12 haplotypes were very similar compared to many other haplotypes (Miller *et al*., 1984).

Phylogenetic analysis based on the conceptual cDNA sequences of the BG genes found in B cells from the four chicken lines as well as the 14 BG genes from B12 haplotype show similar results as in the T cell project, except for one gene, NBBGf. Generally speaking, all the BG genes found in B cells have haemopoietic BG gene features (haemopoietic 5' UTRs) with either type 1 or type 2 cytoplasmic tails, except NBBGf. NBBGf has haemopoietic 5' UTR but is closely related to a tissue BG gene (BG6) from the Ig-V domain to the cytoplasmic tail region, and is clustered with BG9 in 3' UTR. Also, except two clades (BG8-9-12 clade and BG7-11 clade), most of the BG genes look like hybrid genes, with some of them grouped together in one region but grouped into different clusters in other regions.

Second, many alternative splicing transcripts were observed for most BG genes found in B cells, which are due to different usage of alterative splice acceptor sites and intron retention, with the latter one resulting in truncated cytoplasmic tails and soluble proteins. In previous T cell project, intron retentions were all seen in the cytoplasmic tail region leading to truncated cytoplasmic tails. On the contrary, in B cells some BG genes have additional region where intron retention happened, the intron between exon 2 (encoding Ig-V domain) and exon 3

(encoding transmembrane region) retained, leading to soluble BG proteins if such cDNA transcripts are translated. The percentages of soluble protein transcripts vary in different BG genes, and as some BG genes were only found with a few clones, it would be difficult to calculate the percentage. However, all the BG genes found with soluble protein transcripts are listed as the following, with the first number after gene name showing soluble protein transcripts and the second showing total number: line N not detected, line P2a one gene out of three [P2aBBGa(27/59)], line 15I two genes out of five [15iBBGa(3/39) and 15iBBGe(1/1)], and line 6_1 one gene out of eight [6BBGa(5/5)]. Such findings are striking, and should be tested at protein level. If these different cDNA transcripts could be proved at protein level, especially the soluble BG proteins in B cells, it will be a novel finding for BG.

Third, only one dominantly expressed BG gene was found in B cells of each of the four chicken lines, which was not expected. We expected that there would be two different BG genes highly expressed in B cells of each chicken line, as that was what we knew from previous work on B12 haplotype where both BG7 and BG9 were highly expressed. However, there was only one dominantly expressed BG gene observed in each line examined. Among the five dominant BG genes, two belong to BG8-9-12 clade, one belongs to BG7-11 clade, while the other one (P2aBBGc) looked like hybrid gene. P2aBBGc has a clear BG7-11 clade feature in cytoplasmic tail and 3' UTR, but is closer to BG8-9-12 clade in other regions. BG8-9-12 clade differs to BG7-11 clade from the signal sequence region to the 3' UTR region, especially the two potentially functional domains, Ig-V domain and cytoplasmic tail.

Then we compared all the BG cDNA sequences generated from this project to those from our previous work as well as those BG cDNA sequences downloaded from GenBank. Many interesting findings were discovered through nucleotide and amino acid sequence alignments and phylogenetic studies.

Firstly, a few BG cDNA sequences treated as independent transcripts were most likely to be the result of PCR errors, and were corrected. By comparing all the BG cDNA sequences from both B cell and T cell projects, most single nucleotide variation was considered as the result of PCR errors and was corrected. Many cDNA sequences that only appeared in one PCR from one cell type but found in the other cell type were confirmed to be real. During data analysis, we tracked each sequence by noting how many clones it represented, and from how many PCRs. For those sequences that only show up in one PCR, we were cautious and did not use such sequences for crucial analyses. Through the comparison work above, many

suspicious sequences were validated and can be used for future work.

In the mean time, through comparing alternative splicing isoforms of the same BG gene from different cell types, it was found that cDNA transcripts encoding soluble BG proteins were found only in B cells. Also, comparatively, many cDNA transcripts encoding truncated short cytoplasmic tails were found in T cells, while longer truncated cytoplasmic tails were seen more often in B cells. However, the most dominantly expressed cDNA transcripts of the same BG gene from T and B cells were normally found to be identical, apart from one exception from line 6_1. In line 6_1, the dominantly expressed BG gene in T cell is 6TBGa, whose conceptual cDNA sequence is identical to that of BG13 from B12 haplotype. However, of all the transcripts found for 6TBGa, the introns after the first 21 nt exon of cytoplasmic tail were retained, resulting in the very short truncated cytoplasmic tail. Also, all the cDNA transcripts of 6BBGa found in B cells contain the introns between exon 2 (encoding Ig-V domain) and exon 3 (encoding transmembrane region), resulting in the soluble proteins. Currently, there is no clue about why line 6_1 is so different to other lines; one possible explanation is that there were some recombination events in line 6_1, but 6TBGa could not meet the function required in T cells, and thus such special alternative splicing transcripts were selected to de-regulate the function of 6TBGa itself.

Secondly, all the data so far support the idea that the chicken BG haplotypes are consistent with the B haplotype. By comparing our BG cDNA sequences with all the others available from our previous work or online, it helped to answer the question that whether different chickens from the same chicken line (or same B haplotype) have the same BG genes. We first compared the BG genes found from line 15I in this project to these previously found from duodenum of another line 15I chicken, and the nucleotide sequence alignments showed that the BG cDNA sequences found from duodenum all belong to 15iBGa with various alternative splicing in the cytoplasmic tails. Then we compared the cDNA sequences from GenBank, derived from line UCD330 chicken which is B21 haplotype, to our BG sequences from line N which is also B21 haplotype. The results confirmed that four GenBank B21 cDNA sequences were actually the same BG gene in line N, NTBBGa, with different alternative splicing in some of their cytoplasmic tails. Two additional GenBank B21 cDNA sequences were very similar to NBBGf with a few differences in cytoplasmic tails and only one nucleotide difference in the Ig-V domains. There was a GenBank sequence not matching to any BG gene from line N in our project. Either our HU primer might not amplify all the haemopoietic BG

genes, or the GenBank sequence might be a tissue BG gene (as it lacks the 5' UTR sequence information) that our HU primers would not amplify.

Thirdly, the annotation of 15 BG genes in Gaga5.0 (BQ haplotype) and the comparison between BQ and B12 haplotypes provided evidence of swapping whole BG clusters between haplotypes during evolution. Together with the evidence above that BG genes are the same in different chickens from the same haplotype, it explained why six BG genes were found identical between B2 and B12 haplotype, and two BG genes were found identical between BQ and B12 haplotype. There was probably recombination between B2 and B12, and between BQ and B12.

Fourthly, the high polymorphism of BG genes was illustrated by comparison of all BG genes from our project to these free range chickens (Camperos). There was no identical BG gene found (based on exon 2 region) between Camperos chicken and all the BG genes analyzed in this project. However, the phylogenetic tree built on nucleotide sequence of Ig-V domains showed that most Camperos BG genes fell into the known BG clades, perhaps suggesting that BG genes from particular BG clade interact with same receptors/ligands. In the previous T cell project, it has been understood that the variation among BG genes from BG8-9-12 clade is localized to distal loops of Ig-V domains without evidence for selection. However, comparison of the BG8-9-12 clade with BG7-11 clade showed that the variation is mainly on two β strands and the link regions between different strands. More strikingly, all these variable amino acids have their side chains pointing out, strongly suggesting that different BG clades interact with different counterparts.

Fifthly, the similarities and differences of BG genes among different haplotypes may help us understand disease resistance. It is quite interesting that some haplotypes shared similar characters in disease resistance turned out to have the same or similar BG genes. For example BQ and B21 haplotypes were both resistant to MD (Abplanalp *et al.*, 1992) and they share at least one identical BG gene and another very similar BG gene based on our observations. Another research found that B19 was the most susceptible to infectious bronchitis virus (IBV) while B2 was the most resistant haplotype to IBV compared to many other haplotypes (da Silva *et al.*, 2017). Here in our research, nine BG genes were found in both T and B cells in line 6_1 (B2), while only three BG genes were found in line P2a (B19), so maybe the large variety of BG genes helps IBV resistance.

To sum up, identification and analysis of 23 BG genes from four different chicken lines provides enormous information for future studies. However, it is still unclear about the whole evolution story of BG genes among different haplotypes. Genomic typing would give us a more complete picture of BG evolution history. Also, it would be extremely interesting to detect BG expression during certain infections, which would help us understand BG functions much more and potentially provide theoretical support for vaccine design in the future.

Chapter 6

Study of BG-Fc fusion proteins and anti-BG monoclonal antibodies

6.1 Introduction

Previously we have examined the expression of BG genes in T and B cells of four different chicken lines at the mRNA level, and observed many striking findings. For example, there was no identical BG gene found between these four chicken lines, although one line (line 6_1) shares many genes with line CB which had been studied long time ago. The number of expressed BG genes varied among these four chicken lines, ranging from three in line P2a to nine in line 6_1. Phylogenetic analysis revealed that the most dominantly expressed BG genes ('functional alleles') in T cells are conserved to the BG8-9-12-13 clade, while in B cells belong to two clades, the BG8-9-12 clade and the BG7-11 clade. More interestingly, alternative splicing was observed in most BG cDNA sequences, mainly due to intron retentions. When intron retentions occurred, the retained intron introduces an early stop codon, resulting in truncated cytoplasmic tail if the intron is in cytoplasmic tail region, or soluble protein if the intron is in the region between exon 2 (encoding Ig-V domain) and exon 3 (encoding transmembrane region). Strikingly, the alternatively spliced cDNA sequences potentially translated to soluble BG proteins were only found in B cells.

In order to confirm and fully understand these findings observed at cDNA level, much work should be done at protein level. For example, where are BG proteins distributed in tissues, can these alternative splicing transcripts be translated into proteins, especially these transcripts encoding soluble BG proteins only found in B cells, etc. Therefore, in this chapter, all the studies are designed to help answer these questions above. We decided to use B12 haplotype (line CB) as a model because B12 is the best studied haplotype so far for BG research (Salomonsen *et al.*, 2014).

The first step was to develop BG-Fc fusion proteins for all 14 BG genes know in the B12 haplotype, which can be used not only for functional assay, cell staining for ligand searching, Ig-V structure crystallization etc., but also for characterizing BG mAbs. Dr. Jan Salomonsen had created many anti-BG hybridomas, of which only a few were characterized (Salomonsen *et al.*, 1991). With the help of the 14 BG-Fc fusion proteins, all these BG mAbs could be examined by enzyme-linked immunosorbent assay (ELISA) and western blot. Once the specificity of the BG mAbs was determined, tissue distribution and BG function can be studied.

6.2 Materials and methods

6.2.1 BG-Fc SigpIg Plus plasmid construction

The Ig-V domains (exon 2 excluding the first two amino acids) of 14 BG genes from B12 haplotype were cloned separately into vector SigpIg Plus, a kind gift from Prof. John Trowsdale's lab, in order to make soluble BG-Fc fusion proteins. The plasmid construction was first carried by Ms. Valeria Radjabova, by whom seven BG-Fc SigpIg Plus plasmids and 29 glycerol stocks were made. To confirm that all the inheritance from Ms. Radjabova was correctly constructed with right nucleotide sequences, these plasmids were transfected into *E. coli* DH5α competent cells individually, followed by minipreps and sequencing; cells from the glycerol stocks were grown, and minipreps were made and sequenced. In total, eleven clones confirmed as having correct sequences were grown and DNA prepared by midiprep, and grown for glycerol stocks, while another three BG-Fc plasmids (BG3-Fc SigpIg Plus, BG10-Fc SigpIg Plus and BG13-Fc SigpIg Plus) needed to be constructed (Table 6.1).

Table 6.1 The background of BG-Fc SigpIg Plus plasmids and the work done by Chen

	From Radjabova		**Work done by Chen**			
	plasmid	Gly stock	transforming	miniprep	Sequencing result	
BG0-Fc SigpIg Plus	yes		√	√	right	
BG1-Fc SigpIg Plus	yes		√	√	right	
BG2-Fc SigpIg Plus		9 vials		√	right	
BG3-Fc SigpIg Plus	yes		√	√	wrong	Re-clone
BG4-Fc SigpIg Plus	yes		√	√	right	
BG5-Fc SigpIg Plus		8 vials		√	right	
BG6-Fc SigpIg Plus	yes	7 vials		√	right	
BG7-Fc SigpIg Plus		1 vial		√	right	
BG8-Fc SigpIg Plus		1 vial		√	right	
BG9-Fc SigpIg Plus		1 vial		√	right	
BG10-Fc SigpIg Plus						clone
BG11-Fc SigpIg Plus		1 vial		√	right	
BG12-Fc SigpIg Plus	yes		√	√	right	
BG13-Fc SigpIg Plus	yes		√	√	wrong	Re-clone

To clone the Ig-V domain of BG3, BG10 and BG13, three pairs of primers were designed, with HindIII and XhoI restriction site sequences included in the forward primers and reverse primers, respectively (Table 6.2). The PCR template for BG3 Ig-V domain amplification was the BG3-Fc SigpIg Plus plasmid made by Ms. Radjabova described above which missed a nucleotide 'T' at the 5' end of BG3 Ig-V domain sequence. The BG10 and BG13 Ig-V domains were both cloned from cG3 cosmid of chicken line CB (B12 haplotype) (Salomonsen, et al. 2014). Then PCR products were separated by gel electrophoresis, and the specific DNA bands around 330 bp on the gel were cut out for purification. The purified DNA and SigpIg Plus vector were both double-digested by HindIII and XhoI (both from NEB), then purified using ISOLATE II PCR and Gel Kit (Bioline), ligated using T4 DNA Ligase (NEB) and transformed into *E. coli* DH5α competent cells. Ten colonies from each cloning were picked to grow overnight bacterial cultures and minipreps were prepared. The successful insertion of BG Ig-V domain fragments into SigpIg Plus vector was further confirmed by sequencing.

Table 6.2 Primers designed to amplify Ig-V domains of BG3, BG10 and BG13

Primer name	Primer detail
BG3 forward primer	5' TATAAGCTTACAGTGGTAGCACC 3'
BG3 reverse primer	5' CCGCTCGAGATCTGACACCTCCAGC 3'
BG10 forward primer	5' TATAAGCTTACGGTGGTGGCACC 3'
BG10 reverse primer	5' CCGCTCGAGATCTGACACCTCCAGCTC 3'
BG13 forward primer	5' TATAAGCTTACGGTGGTGGCACC 3'
BG13 reverse primer	5' CCGCTCGAGATCTGACACCTCCAG 3'

Note: AAGCTT in red is the HindIII cleavage sequence, and CTCGAG in red is the XhoI cleavage sequence.

6.2.2 Selection of stable cell lines expressing BG-Fc fusion proteins

The 14 BG-Fc SigpIg Plus plasmids were transfected into HEK293 cells (a stock vial through Ms. Radjabova from Prof. John Trawsdale's lab) in 6-well plates separately using jetPEI in vitro DNA kit (Polyplus Transfection) following the manufacture's protocol. After 48 h of

transfection, cells of each well were transferred into individual 25 cm^2 flasks with 5 mL DMEM (containing 10% FBS), and 100 µl 50 mg/mL Geneticin 418 (G418 for short) (Santa Cruz Biotechnology) was added into the culture medium to select the cells with successful transfection. During the selection, most cells died in the first week, and only a few cells survived. To keep these surviving cells expanding, old culture medium was replaced by fresh medium with G418 once per week. On average, one to two months would be taken for the cells to become confluent. Stable cell lines were frozen in vials with freezing medium (90% FBS and 10% DSMO) overnight at -80°C and then transferred into a liquid nitrogen tank. During the whole process, the BG-Fc fusion proteins were checked regularly by SDS-PAGE and western blot using goat anti-human IgG-HRP antibody (AbD Serotec) described below.

6.2.3 Checking the fusion protein expression by SDS-PAGE and western blot

To test the expression of BG-Fc fusion proteins, 10 µl of the cell culture supernatant was collected and incubated at 75°C for 10 minutes with 2.5 µl NuPAGE LDS sample buffer (Life Technology) which was freshly mixed with dithiothreitol (DTT) (Melford Laboratories) to make the final concentration of 100 mM for DTT. The heated proteins were separated by SDS-PAGE (10% separating gel); then the gel was semi-dry transferred onto Immobilon-P Membrane (Merck Millipore) using Trans-Blot SD Semi-Dry Transfer Cell (Bio Rad). The membrane was blocked at room temperature for 1 h or at 4°C overnight using TBST with 5% skim-milk. After being washed 3 times (5 min each) with TBST, the membrane was incubated with goat anti-human IgG-HRP antibody (AbD Serotec) diluted at 1:5000 in TBST with 5% BSA at room temperature for 1 h. After incubation, the membrane was washed for 4 times (8 min each) with TBST, and then was incubated with the chemiluminescent detection reagent ECL (GE Healthcare, Life technology) and exposed using X-ray Film (Konica Minolta).

6.2.4 BG-Fc fusion protein purification

In order to reduce the Ig contamination coming from cell culture medium, the medium was replaced with serum free DMEM when BG-Fc stable cell lines were grown to 80% competency. Three to five days after changing medium (depending on the cell growth, if more than half of the cells died, the supernatants were collected immediately), cell culture supernatants were collected and concentrated using Vivaspin® concentrator 10 kDa (Sartorius) by centrifuging at 3000-5000 g for 10-25 min. Then the concentrated solutions

were washed three times, each time adding 2 mL of protein washing buffer (PWB, 0.1 M sodium phosphate, 0.15 M NaCl, pH 7.4) into the Vivaspin tube and repeating the concentrating procedure above. Finally, the proteins were concentrated to 500 µl PWB and transferred into an EconoSpin® mini spin column, to which had been added with 20 µl protein G agarose beads solution (2.5 mg/ml, Sigma®). The column was then placed on a slow rotator for 1 hour at room temperature to allow complete binding between proteins and resin. Then the column was washed three times with 700 µl PWB by gently inverting the column up and down, and centrifuging at 6,500 g for 1 min, and finally centrifuged at 6,500 g for 1 min to ensure no PWB was left in the column. Then 80 µl protein elution buffer (0.2 M glycine-HCl, pH 2.7) was added to the column and incubated with the protein-resin mixture at room temperature for 5 min, and finally the proteins were collected by centrifuging at 8,000 g for 1 min. To avoid protein degradation, 1 µl 0.1 mM 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF) (Sigma) was added and the proteins were kept at -20°C for immediate use or -80°C for longer storage.

6.2.5 BG hybridoma supernatant

A total of 290 flasks of BG monoclonal antibody (mAb) supernatants were made by Dr. Jan Salomonsen about 30 years ago. Some of these mAbs had been used before to immunoprecipitate BG antigens (Salomonsen *et al.*, 1991); the records regarding to the detailed background of these mAbs were missing; however, the names of hybridomas which these supernatants belong to are marked on the flasks. For the convenience of recording, and to ensure the following experiments were attributable, all the flasks were renamed and reorganized from BGmAb1 to BGmAb290.

6.2.6 Development of a sandwich ELISA

In order to determine the specificities of each BG mAb, a sandwich ELISA was developed to detect the interactions between BG-Fc fusion proteins and the BG mAbs. The general procedure for a sandwich ELISA has five steps: step 1, the plate is coated with a capture antibody (goat anti-mouse IgG in this chapter); step 2, antigens (BG mAbs) that binds to capture antibody are added; step 3, samples (BG-Fc fusion proteins) that may bind to the antigens are added; step 4, the detecting antibody, enzyme-linked secondary antibody (anti-human IgG Fc-HRP) is added; step 5, substrate is added and colour is detected by ELISA plate reader.

A sandwich ELISA procedure was established through two stages. In stage one, a standard ELISA was developed and optimized by using standard antigen and antibody. Compared to conventional ELISA, more procedures and reagents are required for the sandwich ELISA, therefore, it is comparatively more difficult to optimize each step above, especially the concentrations of each critical reagents, coating and incubating time etc.. In order to avoid uncertainties during the experiment, standard antigen (supernatant of stable cell line expressing B7H6-Fc fusion protein with accurate measurement of the concentration of B7H6-Fc fusion proteins) and antibody (mAb CH31, supernatant of a mAb hybridoma against B7H6 with accurate measurement of its concentration), kind gifts from Dr. Chiwen Chang, were used to set up the basic sandwich ELISA protocol. In stage two, the protocol was adjusted accordingly when switching to our own antigens (BG-Fc fusion proteins) and antibodies (BG mAbs). Finally, an optimized protocol for BG mAb-sandwich ELISA was set up, and the procedure is briefly described as below.

a. Coat the plate with 1.25 µg goat anti-mouse IgG, IgA, IgM (H+L) (BioFX) overnight at room temperature (RT);

b. Block the plate overnight at RT using blocking buffer (1% BSA in PBS, pH7.2-7.4, 0.2 µm filtered);

c. Incubate the plate with 80 µl BG mAb supernatant for 30 min at RT;

d. Wash the plate three times for 5 min each time (3x5 min) with TBST [140 mM NaCl, 50 mM TrisCl, 0.05% Tween-20 (pH 8.0)] followed by one more wash with PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM K_2HPO_4);

e. Incubate the plate with 80 µl BG-Fc supernatant for three hours at RT;

f. Wash the plate with TBST (3x5 min) followed by one more wash with PBS;

g. Incubate the plate with 80 µl detecting antibody [1 µl mouse anti-human IgG Fc-HRP diluted in 900 µl blocking buffer with 100 µl mouse serum (ThermoFisher Scientific)] for 30 min at RT;

h. Wash the plate with TBST (3x5 min) and another three times with PBS;

i. Incubate the plate with 80 µl substrate A (stabilized hydrogen peroxide) and B

(stabilized tetramethylbenzidine) mixture (1:1) (R&D Systems) for 5 min at RT;

j. End the reaction by adding 40 µl stop solution (3M HCl);

k. Read optical densities at 450 nm (OD_{450}), considering the ones with OD_{450} more than 0.5 as positive.

6.2.7 SDS-PAGE and western blot to determine the BG mAbs specificity

In this project, two protocols of doing SDS-PAGE and western blot were used. The first protocol was very similar to the one described in section 6.2.3, except for two things. First, during the SDS-PAGE, the purified BG-Fc fusion proteins (300 ng/each protein diluted in PBS), instead of BG-Fc fusion protein supernatants, were loaded onto the gel. Second, during the western blot procedure, the membrane with transferred BG-Fc fusion proteins was firstly incubated with BG mAbs, washed, then incubated with detecting antibody, the goat anti-mouse IgG-HRP, washed again, and finally incubated with ECL for X-ray film exposure.

The second protocol was a homemade method to use the first protocol at a large scale; since there were many flasks of BG mAbs supernatant to screen with 15 different proteins (14 BG-Fc fusion proteins and one Fc fusion protein as negative control), it would be extremely time consuming and costly using the conventional protocol. As indicated in figure 6.1, six major steps were performed. Step 1, the SDS PAGE gel (reduced or non-reduced gel) was prepared with the special comb which yielded a small well and a large well, and the standard protein marker was loaded in the small well while the BG-Fc fusion protein (300 ng dissolved in 30 µl PBS) was loaded into the large well. Step 2, after completing the SDS-PAGE, the gel was transferred onto the Immobilon-P membrane, a small portion of the membrane covering standard protein marker and BG-Fc fusion protein was cut to complete the western blot in order to help identify the location of BG-Fc fusion protein on the membrane. Step 3, the rest of the membrane was cut into narrow strips, and the strips were reorganized following the order of 'BG0-Fc, BG1-Fc, ..., BG13-Fc' and held by tape to form a homemade 14 BG-Fc strip tape. Step 4, the homemade strip tape was incubated with a particular BG mAb supernatant with its ID number labelled on the tape. Step 5, to avoid contamination, all the following procedures including washing and incubating the detecting Abs were performed in their own small containers. Step 6, nine or 12 homemade strip tapes were placed on the hard plastic film and added with ECL for X-ray film exposure.

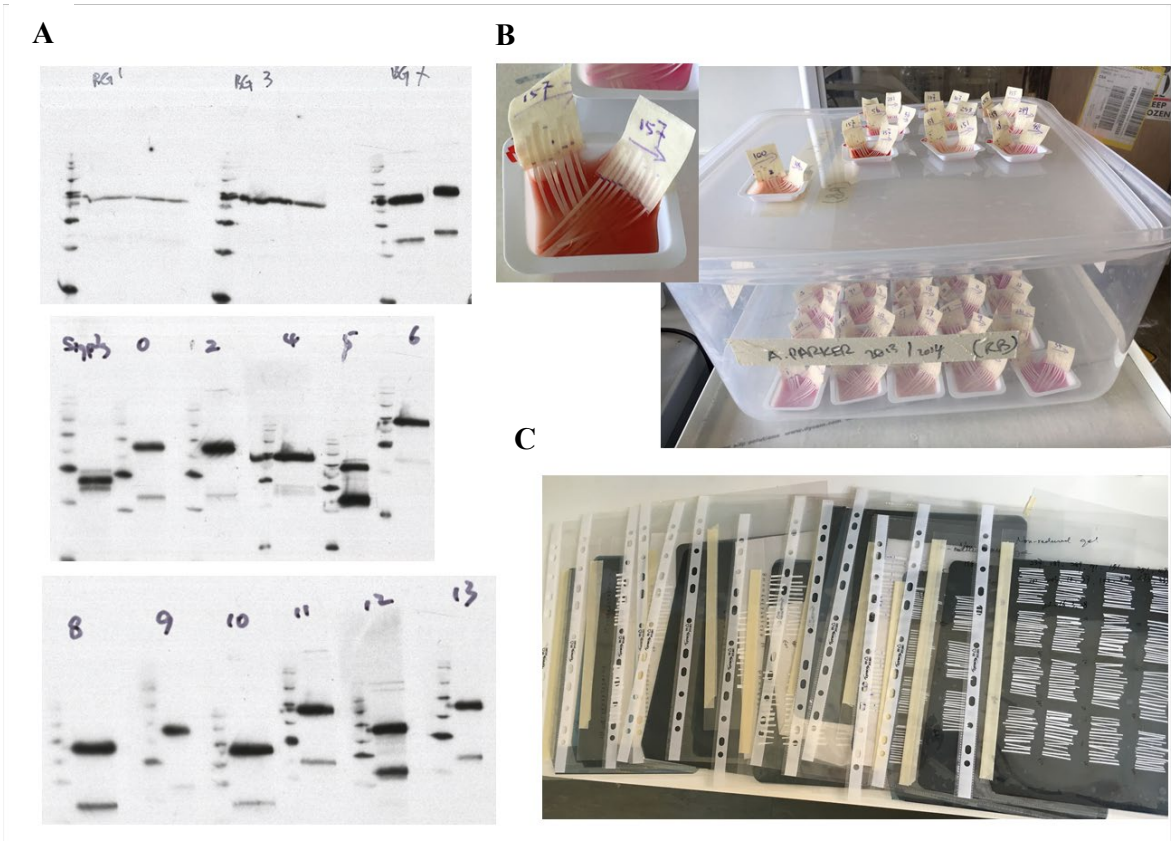


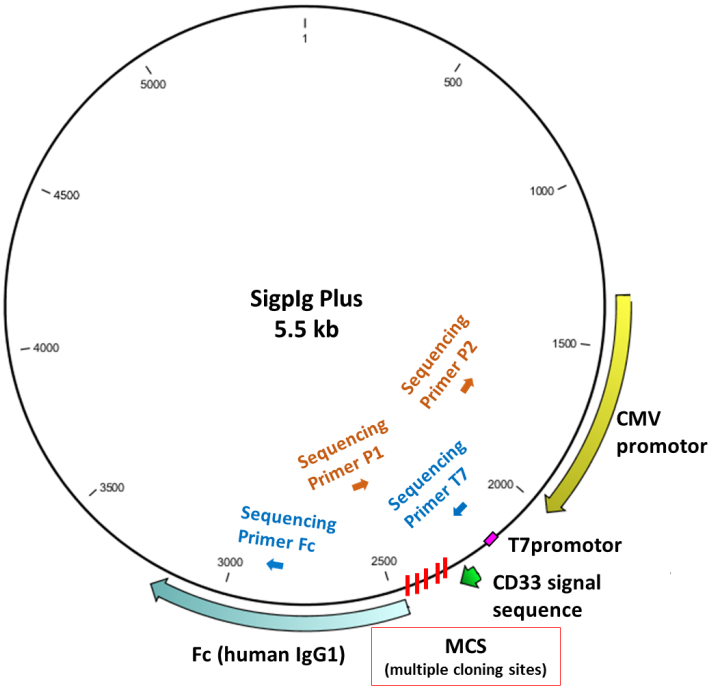
Figure 6.1 Homemade method for handling large number of western blots.

A, fourteen BG-Fc fusion protein and Fc protein (labelled with SigpIg on membrane) were run by SDS-PAGE (reduced) gel separately, then gels were transferred onto Immobilon-P membrane (membrane), and a small portion of each membrane (including the entire area of standard protein marker lane) was cut and incubated with the goat anti-mouse IgG-HRP followed by X-ray film exposure. B, the homemade strip tapes each containing 15 strips (including 14 BG-FC fusion proteins and one Fc protein) are incubated with 2 mL BG mAb supernatants in small container boxes. C, nine or 12 homemade strip tapes were laid on one plastic film for X-ray exposure.

6.3 Results

6.3.1 The design of BG-Fc fusion proteins

The SigpIg Plus plasmid has been used to generate fusion proteins for functional studies in Prof. John Trowsdale's lab for a long time (Smith *et al.*, 2010; Ammann *et al.*, 2012; Boyle *et al.*, 2013). The detailed sequence information of the whole plasmid was missing. However, a region of roughly 2 kb upstream and downstream of the multiple cloning site (MCS) was sequenced by myself in order to fully understand how this plasmid works. As shown in figure 6.2, the empty vector was firstly sequenced by sequencing primer T7, then sequencing primer Fc and P1 were designed based on the first sequencing result and were used for sequencing, finally sequencing primer P2 was designed based on previous P1 sequencing result and applied for sequencing. All the oligo information for sequencing primers is: T7 (5' TAATACGACTCACTATAGGG 3'), P1 (5' CCGCTCTCGAGATCTGACAC 3'), Fc (5' TGAGCCACGAAGACCCTGAGG 3'), P2 (5' CAATAGGGGGCGTACTTGGC 3'). The empty SigpIg Plus vector is about 5.5 kb according to its band size on DNA gel electrophoresis after double enzyme digestion; upstream of the MCS, there is a cytomegalovirus (CMV) promoter, followed by the CD33 signal sequence; and downstream of the MCS is the human IgG1 Fc region. The restriction enzymes chosen for constructing BG Ig-V sequence into SigpIg Plus vector were HindIII and XhoI.



CD33 signal sequence

HindIII

ATGCCGCTGCTGCTACTGCTGCCCCTGCTGTGGGCAGGGGCCCTGGCTATGGATAAGCTTTG

XhoI

GTGGAGCTGGAGGTGTCAGATCTCGAGAGCGGAGGAAGCAGCGGAGGAAGCAGCGGAG

EcoRV XbaI BamHI

GAAGCAGCGGAGGAAGCGATATCTCTAGACCAGTGTGCTGGGCGGCCGCGGATCCCATCG

Fc (human IgG1)

AAGGTCGTGGTGGTGGTGGTGGTGATCCCAAATCTTGTGACAAAACCTCACACATGCCCACT
GTGCCCAGCACCTCCAGTCGCGGGACCGTCAGTCTTCCTCTTCCCCCCAAAACCCAAGGACA
CCCTCATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTGGTGGACGTGAGCCACGAAGA
CCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAA
GCCGCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCAC
CAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGGCCTCCCATCCT
CCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGTGTACACCCT
GCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGTCAGCCTGACCTGCCTAGTCAAAGGC
TTCTATCCCAGCGACATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTAC
AAGGCCACGCCTCCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGT
GGACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCT
GCACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGA

Figure 6.2 SigpIg Plus vector and ORF encoding BG-Fc fusion protein.

CD33 signal sequence is in dark green, followed by multiple cloning sites coloured in grey with multiple restriction enzyme sites underlined and in red or purple, then followed by Fc sequence which is a fragment of human IgG1. The locations where the sequencing primers (T7, Fc, P1 and P2) designed for sequencing SigpIg Plus vector are indicated.

The mature Ig-V domain of BG genes consists of 113 amino acids, but during the initial BG-Fc SigpIg Plus plasmid construction done by Ms Radjabova, the first two amino acids were not included. To keep the work consistent, all the BG-Fc SigpIg Plus plasmid constructions done by myself did not include the first two amino acids either. Therefore, a secreted BG-Fc fusion protein should have 384 amino acids in total (1152 nucleotides with 333 bp encoding BG Ig-V domain, 123 bp as a linker peptide and 696 bp encoding human IgG1 Fc fragment). Secreted Fc fusion proteins are known to be dimers (Smith *et al.*, 2010; Lobner *et al.*, 2017); the BG-Fc fusion proteins should be dimers too, as lots of cysteines are observed in Fc region which could form disulfide bonds to link two Fc fragments. Also, according to previous structure modeling work on BG1 (Chattaway *et al.*, 2016), the cysteines showed as yellow sticks and surface labeling on the two monomers would form a disulfide bond and hold the two Ig-V molecules as a dimer (Figure 6.3).

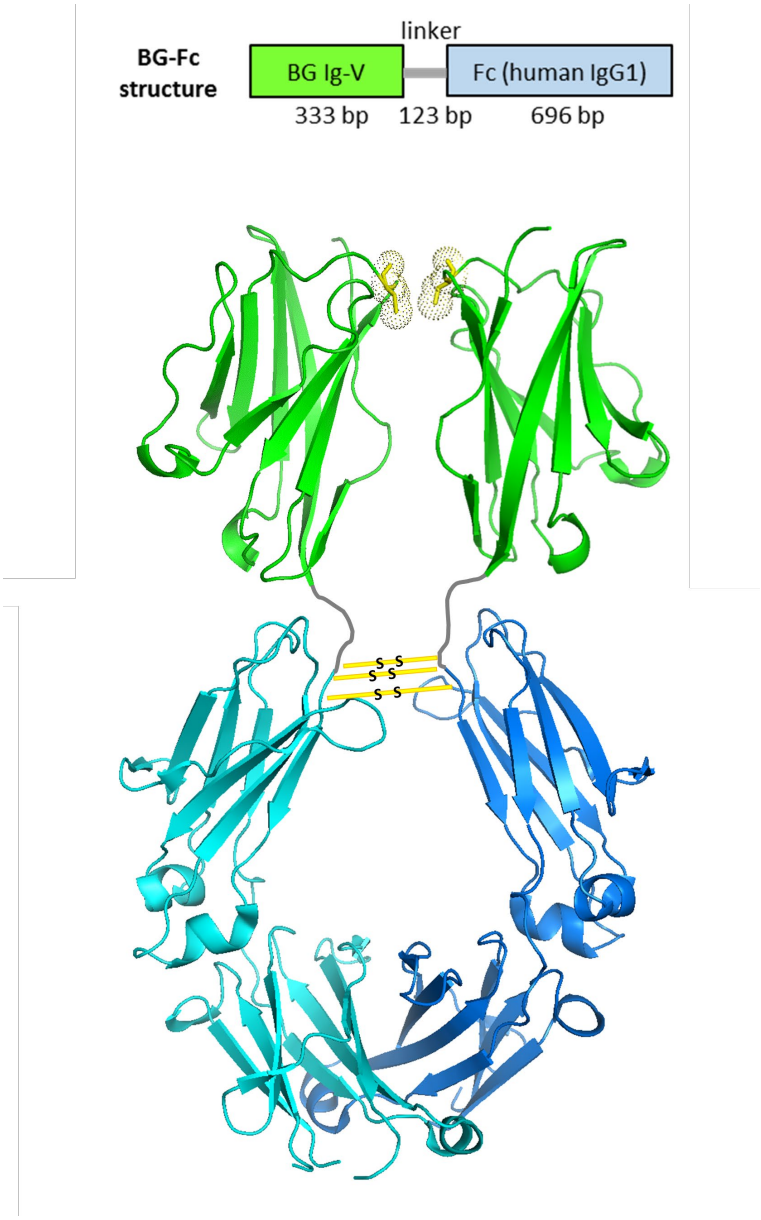


Figure 6.3 Illustration of the BG-Fc fusion protein with linear structure on the top and modeling dimer structure at the bottom. BG Ig-V domain in green, linker peptide in grey and Fc domain in blue. Cysteines that potentially form a disulfide bond are highlighted with stick and surface on the Ig-V structure; three pairs of cysteines forming disulfide bonds in the linker peptide are labeled with yellow sticks. The model for Ig-V domain of BG8 was built by Swiss-Model (https://swissmodel.expasy.org/) based on the template of the MOG molecule (PDB ID 3csp.1) sharing 40.35% identity in amino acid sequences. The Fc domain was from the crystal structure of human IgG1-Fc homodimer (PDB ID 5JIK) (Lobner *et al.*, 2017).

6.3.2 Fourteen Ig-V domains of BG genes of B12 haplotype were successfully cloned into SigpIg Plus vector

Nucleotide sequence alignments between all the constructed BG-Fc SigpIg Plus plasmids and the corresponding Ig-V domains from all 14 BG genes of B12 haplotype showed that the 14 BG Ig-V sequences were successfully inserted into the SigpIg Plus vector, and all the 14 BG-Fc SigpIg Plus plasmids were successfully constructed. As illustrated in figure 6.4, the first two amino acids in the Ig-V domains of mature BG proteins (QL/S/N) were not cloned into the plasmid, which instead were replaced by the two amino acids (KL) encoded by the HindIII sequence. According to SigalP4.1 (http://www.cbs.dtu.dk/services/SignalP/), the signal peptide cleavage site prediction tool, the mature protein should start at the 17^{th} amino acid (M) of the CD33 signal peptide region. Therefore, in theory, fourteen secreted BG-Fc fusion monomers should all start with 'MDKL' at its N-terminal, then the specific BG Ig-V domain of 113 AA, followed by the same linker peptide (61 AA) and the same Fc fragment (232 AA).

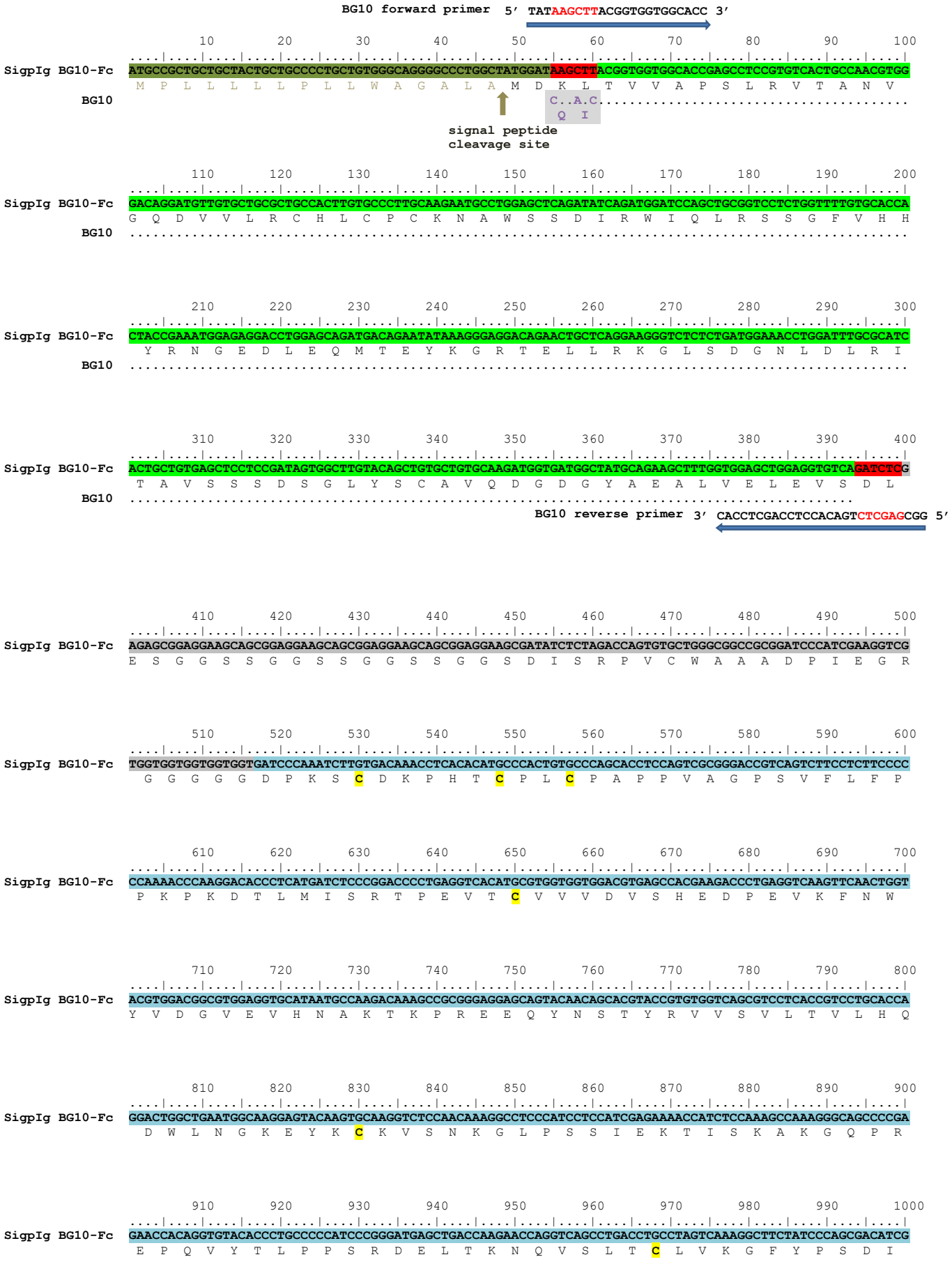


Figure legend is shown on the page following the figure

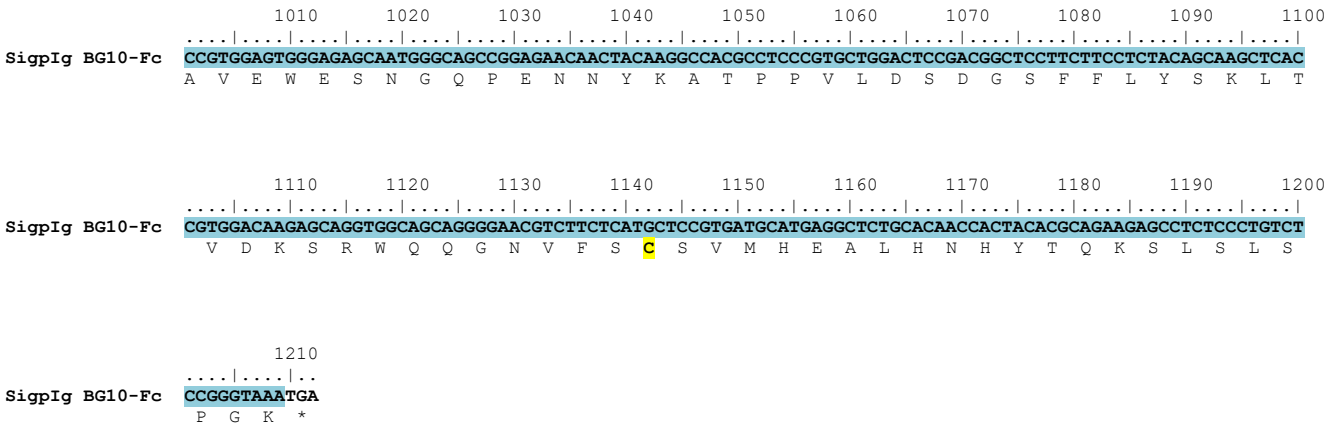


Figure 6.4 An example of nucleotide sequence alignment between the constructed SigpIg BG10-Fc plasmid and cDNA sequence encoding mature Ig-V domain of BG10. CD33 signal sequence is labelled in dark green, followed by HindIII cleavage sequence in red, then Ig-V sequence of BG gene, XhoI cleavage sequence, linker sequence and ending with human Fc sequence. The location of primers to clone Ig-V region of BG10 are illustrated with the HindIII and XholI cleavage sequence coloured in red. The signal peptide cleavage site is indicated between the 16th and 17th amino acids.

6.3.3 Fourteen stable cell lines expressing BG-Fc fusion proteins were selected

Fourteen stable cell lines expressing individual BG-Fc fusion proteins were made in order to produce large amounts of BG-Fc fusion proteins for characterization of BG proteins, as well as for other studies in the future. Fourteen BG-Fc SigpIg Plus plasmids were transfected into HEK293 cells individually under G418 selection and the selection procedure took about two to three months before the cells become stable. During the process, supernatants of cell cultures were collected frequently to test protein expression by SDS-PAGE (reduced gel) followed by western blot with goat anti-human IgG-HRP antibody. Theoretically, the size of BG-Fc fusion proteins should be around 43 kDa calculated by bioinformatics tool DNASTAR Lasergene Suite (DNASTAR).

The stable cell lines producing large quantities of BG-Fc fusion proteins were expanded, aliquoted and frozen down. However, some cell lines didn't yield satisfactory amounts of fusion proteins, so new rounds of transfection and selection were carried out until appropriate stable cell lines were made. As shown in figure 6.5, not all stable cell lines expressed BG-Fc fusion proteins with equally high concentration. However, they all met the requirements for our experiments in this project.

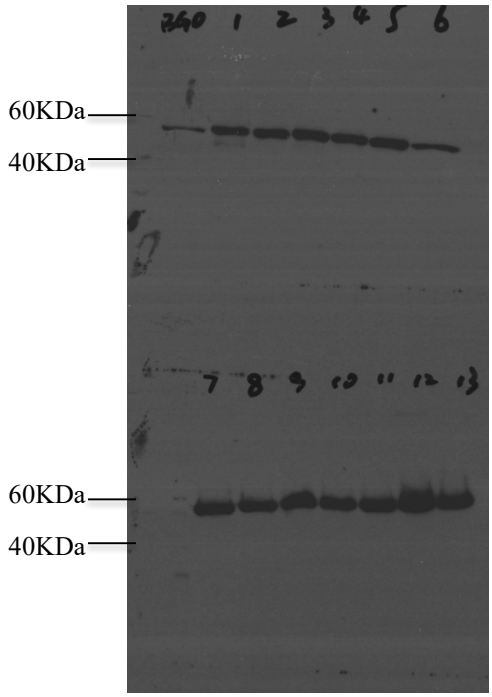


Figure 6.5 Western blot of 14 BG-Fc fusion proteins expressed by BG-Fc stable cell lines. Western blot to show goat anti-human IgG–HRP staining of protein from 14 BG-Fc stable cell lines. X-ray film was exposed for less than 5 seconds. Lanes 0 to 13 are representatives of protein from supernatant of BG0-Fc stable cell line to BG13-Fc stable cell line.

6.3.4 A method to quickly screen a large number of BG mAbs by sandwich ELISA was developed

ELISA is a highly sensitive, versatile and quantitative method for detecting various forms of antigens (Chang, 2004; Chan, 2013). Sandwich ELISA has a higher specificity compared to other types of ELISA methods. It provides a better solution to meet the aim of this project: to find very specific mAbs which can distinguish as many different BG proteins as possible. However, given how complicated the sandwich ELISA procedure is, many critical elements could easily influence the results (for example, the choice of plate, coating and detecting antibody, blocking buffer etc.). With much help from Dr. Chiwen Chang, a sandwich ELISA method was established and successfully used to screen the interactions between all 290 flasks of BG mAb supernatants with all 14 different BG-Fc fusion proteins. Two important decisions contributed to this success.

First of all, a standard antigen, B7H6-Fc fusion protein was used to establish and optimize the sandwich ELISA protocol. B7H6-Fc fusion proteins were expressed from a HEK293 stable cell line using the same culture medium as our BG-Fc stable cell lines, which meant that the background of B7H6-Fc supernatant was very similar to that of BG-Fc supernatant. The B7H6-Fc fusion protein has been characterized very well and been used to develop a specific mAb, CH31 (Dr. Chiwen Chang, personal communication). In comparison, the concentration of BG-Fc fusion proteins in our BG-Fc fusion protein tissue culture supernatants had not been measured at that time, and the BG mAb supernatants had not been characterized by the 14 BG genes from B12 haplotype. If we directly used our antigens and antibodies to establish and optimize the sandwich ELISA protocol, there would have been much uncertainty and confusion. Therefore, we used B7H6-Fc fusion protein and mAb CH31 to test the protocol, which quickly helped us to set the right parametres for the assay and greatly simplified the process.

Second, two rounds of sandwich ELISA were applied to test all 290 BG mAbs. The aim of this experiment was to examine which BG mAb specifically recognize which BG-Fc fusion protein. With 290 hybridoma supernatants and 14 BG-Fc fusion proteins, plus triplicates for each reaction (290x14x3=12180, roughly 130 96-well plates), it will make the task extremely tedious and costly. Eventually, a solution of two rounds of screening was created. As demonstrated in figure 6.6, during the first round, all 14 BG-Fc fusion protein supernatants were mixed as the antigen to screen the 290 BG mAbs, and only those mAbs with positive

results in the first round experiment were tested in the second round sandwich ELISA. In the second round, each BG-Fc fusion proteins were tested separately. To test if this solution works or not, B7H6-Fc fusion protein supernatant, which yields similar amounts of fusion protein as most BG-Fc fusion protein supernatants do in the western blot, was diluted using SigpIg Plus supernatant (stable cell line expressing only the Fc protein). As shown in figure 6.7 (A), there was no obvious influence on the OD values between dilution of B7H6-Fc supernatant from 1:2 to 1:16, indicating 1:14 dilution of BG-Fc supernatant was sufficient enough to be detected in our sandwich ELISA test. To confirm that this method worked with our samples, a few BG mAbs which have been used in previous research (Salomonsen *et al.*, 1991) were characterized using all the 14 BG-Fc supernatants separately, and then tested again using the mixture of all 14 BG-Fc supernatant together. The results showed that mixing the 14 BG-Fc supernatants did not influence the results, as illustrated by one example in figure 6.7 (B) and (C).

Also it is worth mentioning that, to ensure the sandwich ELISA we developed was good enough, one repeat of all the tests was done using a commercial kit, goat anti-mouse IgG microplate (R&D). The only concern of using this kit was that the pre-coated antibody in the commercial kit was directed against IgG, while we did not know the exact Ig types of most of our BG mAbs since the background information was missing. The results showed no difference between our homemade kits compared to the commercial kits (although our homemade kits had a higher background, this did not influence the positive/negative judgement). B7H6-Fc tissue culture supernatants and mAb CH31 were used as positive control to ensure the sandwich ELISA kits working. B7H6-Fc tissue culture supernatants alone were used as negative control when testing BG mAbs reactions with BG-Fc tissue culture supernatants.

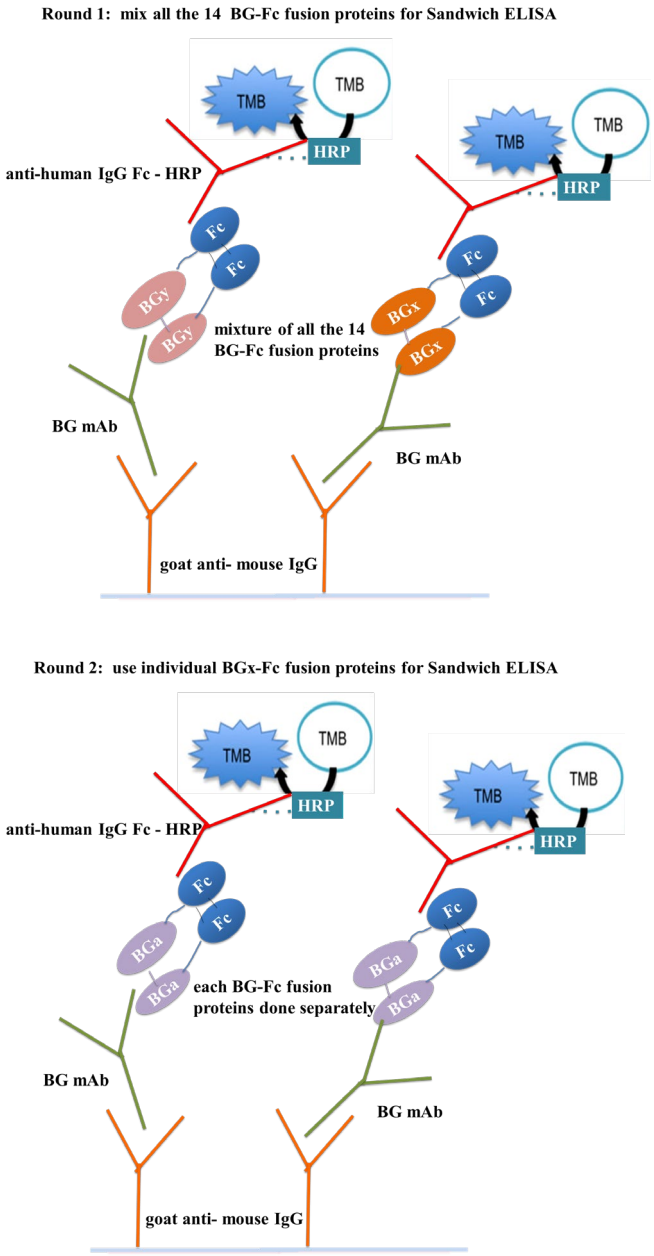


Figure 6.6 Demonstrations of two rounds sandwich ELISA procedure to screen all 290 BG hybridoma. In round 1, the 96 well plate was first coated with goat anti-mouse IgG (coating antigen); each BG mAb was added into a separate well to bind with the coating antigens and was thus immobilized into the well; then the mixture of all 14 BG-Fc fusion protein supernatants were added into each well, and the ones which recognized particular BG mAbs from these wells would be conjugated, otherwise they would be washed off; finally all anti-human IgG Fc-HRP antibodies were added into each well for the ECL exposure. All the BG mAbs showing positive results were tested further in round 2 sandwich ELISA, which only differed in one aspect compared to round 1: the individual BG-Fc fusion protein supernatant was added instead of the mixture of all 14 BG-Fc fusion protein supernatants.

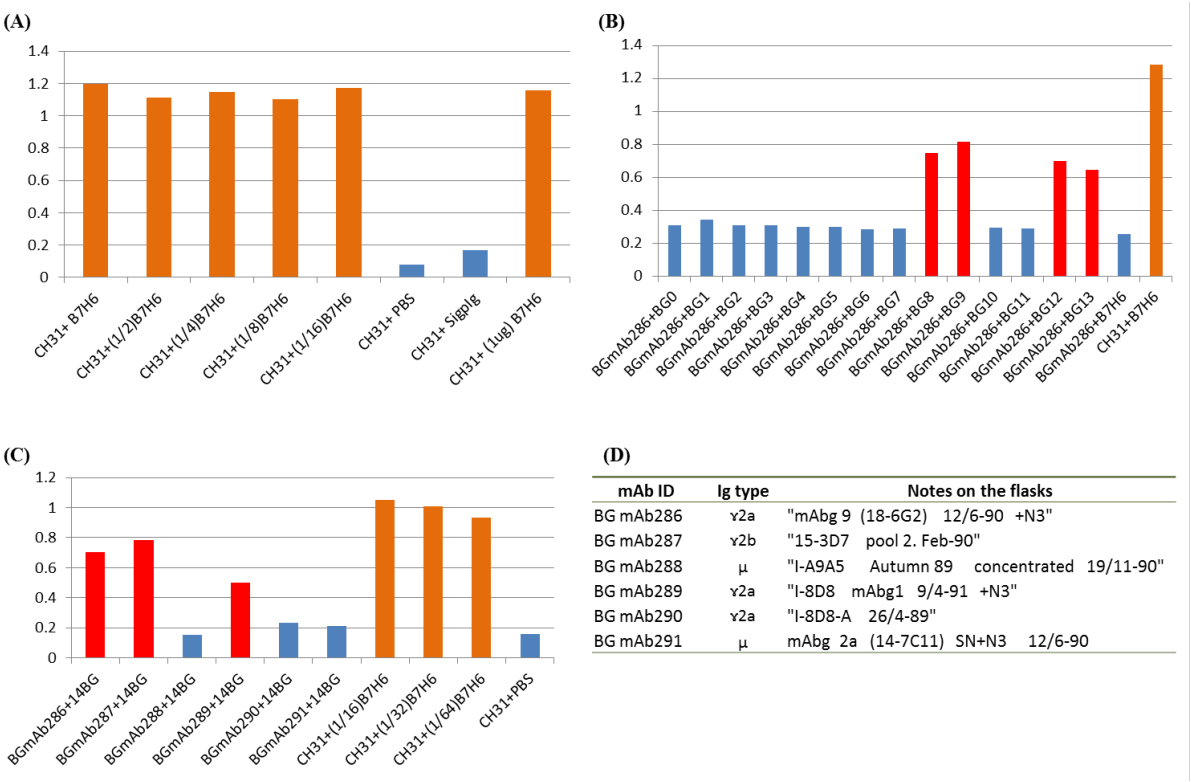


Figure 6.7 Examples of sandwich ELISA results. (A). Serial dilution of B7H6-Fc fusion protein supernatant (from 1:2 to 1:16) didn't influence the interactions between mAb CH31 and B7H6-Fc fusion protein. (B). BG mAb286 recognized BG8, 9, 12 and 13. (C). BG mAb286 showed positive results when all 14 BG-Fc fusion protein supernatants are mixed. (D). The labels on the flasks of BG mAbs used in (C), which isotype determinates from Salomonsen *et al.*, 1991.

6.3.5 Sixty three BG mAbs out of 290 recognize BG-Fc fusion protein with interesting patterns by sandwich ELISA

The two rounds of sandwich ELISA strategy described above worked well and in total 63 flasks of BG mAb supernatants out of 290 were found to specifically recognize certain BG-Fc fusion proteins. As shown in table 6.3, according to the limited information labeled on the flasks, some of these supernatants are most likely derived from the same hybridoma, as they were marked with same labels and the ELISA screening showed almost the same results. For example, mAb229 and mAb158 were both marked with '14-7C11', and they both recognized BG4-Fc and BG5-Fc fusion proteins (BG4 and BG5 for short), with mAb229 having weak reaction to BG2, while mAb158 having weak reaction to BG2 and BG3; mAb195, mAb199, mAb200, mAb201 and mAb236 were all labeled with 'I-2C10', and they all recognized BG8 and BG9 with weak reactions to BG12 and BG13.

Table 6.3 The records of 63 BG mAb supernatants recognizing BG-Fc fusion proteins

mAb No.	Result	First screening	Second screening	Third screening	Forth screening	notes on the flask/tube
229	4,5, (2 weak)	4,5, (2,3 weak)	4,5, (2 weak)	4,5, (2 weak)		14 7C11 SN+N3 28.4.89
158	4,5 (2,3 weak)	4,5 (2,3 weak)				14-7C11 pool Sept. 91. +N3
110	8,9	8,9				2C-10-2 7/2
137	8,9	8,9				II 301 31/3 concentrated ???pool 22/7 91 +N3
148	8,9	8,9				II 349 14/4 89 concentrated + ????? 18/7 91 added N3
153	8,9	8,9,(1,3 weak)	8,9	8,9		Fu II 349 4/5.89 +0.5%N3
195	8,9,(12,13 weak)	8,9,(12,13 weak)				I 2C.1D 28/12 87 1/11.90 concentrated
199	8,9,(12,13 weak)	8,9,(12,13 weak)				I-2C10-1 7.2.
200	8,9,(12,13 weak)	8,9,(12,13 weak)				I2C10 20.11. SN + N3
201	8,9,(12,13 weak)	8,9,(12,13 weak)				I 2C10 30/10 concentrated 19/11.90
236	8,9,(12,13 weak)	8,9,(12,13 weak)				I 2C10 +N3 9/4 91
128	1,6,13	1,6,13				I 18-D.11 (28/2.90) concentrated 30/10.90
129	1,6,13	1,6,13				I 18D11 30/10 concentrated 16/11.90
130	1,6,13	1,6,13				I 18D11 20/11. concentrated pool 20/11 90
8	8,9,12	8,9,12				II-427 + N3 10/10 90
15	8,9,12	8,9,12				II-240 + N3 10/10 90
55	8,9,13	8,9,13				I8D8-3 15.5 pool
57	8,9,13	8,9,13				I8D8-3 pool 1 Feb 90
75	8,9,13	8,9,13				J8D8 A 8.5.89 (filtered 24/10/90 due to ?????? Growth by Fiona
100	8,9,13	8,9,13				I -2C10 2/4.91
127	8,9,13	8,9,13				I - 18C4-3 7.2.
131	8,9,13	8,9,13				I 8D8-33 28/12 89 concentrated pool 6/11 90
193	8,9,13	8,9,13				I-1A8-2 7.2.
194	8,9,13	8,9,13				I-2E3 7.2.
237	8,9,13	8,9,13				I-2E3-2 7.2.
245	8,9,13	8,9,13				I-18C4-4 7.2.
189	8,9,13 (**B**)	8,9,13 (**B**)				? F16g 9. (18-6G2) +N3 18/6.90
196	8,9,13 (**B**)	8,9,13 (**B**)				mAbg 9 +0.1%N3 (18-6G2) 25/6-90
235	8,9,13 (**B**)	8,9,13 (**B**)				I-1A8-1 7.2.
56	8,9,13 (1,12 weak)	8,9,13 (1,12 weak)	8,9,13	1,6,9	8,9,13 (1,12 weak)	I8D8-3 pool 1 Feb 90
9	8,9,12,13	8,9,12,13				II-409 + N3 10/10 90
286	8,9,12,13	8,9,12,13				mAbg 9 (18-6G2)
289	8,9,12,13	8,9,12,13				I-8D8 mAbg1 9/4
91	8,9,12,13(**B**)	8,9,12,13 (B)				I-18H6 9.2.
123	1,8,9,12,13	1,8,9	8,9,12,13	almost recognize all	1,8,9,12,13	I 17A8-1 16/689 concentrated 20/11.90
151	1,8,9,12,13	1,8,9,13 (**B**)	1,8,9,12,13 (**B**)	1,8,9,13 (**B**)	1,8,9,12,13	FII8 II 431 25.5.85
251	1,8,9,12,13	8,9 (**B**)	9,13 (**B**)	1,9,13 (**B**)	1,8,9,12,13	I-18H6-2 7.2.
47	2,5,7,11	2,5,7,11				15 - 3D7 pool 1 Lab 80 95
49	2,5,7,11	2,5,7,11				15 - 3D7 pool 1 Feb 80
176	2,5,7,11,(1 weak)	2,5,7,11,(1 weak)				15-3D7 SN +0.1%N3 13/3.90
287	2,5,7,11?	11 (1 weak)	11 (1 weak)	5,7,(2 weak)	2,5,7,11	15-3D7 pool 2. Feb
190	4,8,9,13	4,8,9,13				18-6G2 1/6-90 +0.1%N3
202	4,8,9,13	4,8,9,13				18-6G2 pool. May 90 +0.1%N3
242	4,8,9,13	1,4,8,9,13	4,8,9,13	4,8,9,13		I-17B8 21/4.89
243	4,8,9,13	4,8,9,13 (1,12 weak)	4,8,9,13	4,8,9,13		I 17B8-11 SN+N3 30.6.83
10	4,8,9,12	4,8,9,12	8,9,12	8,9		II-390 + N3 8/10 90
37	4,8,9,12,13	4,8,9,12,13				I 17B8-2 27.5 Pool
38	4,8,9,12,13	4,8,9,12,13				I 17B8-1 19.5 Pool
39	4,8,9,12,13	4,8,9,12,13				(I) 17B8 12.5.89
231	1,4,8,9,12,13	1,4,8,9,13	4,8,9,13	1,4,8,9,12,13	1,4,8,9,12,13	I8IE10 SN+N3 28.6.89
107	5	5 (1 weak?)	5	5		14-8B5 3/1-89
154	5	5 (**B**)	5	5		14-8B5 3/1-89
157	5	5	5	5		148B-5 30/10 concentrated 18/11.90
155	5 (**B**)	5 (**B**)				148B5 20.11 SN+N3
156	5 (**B**)	5 (**B**)				14-8B5 28/12.89 concentrated 31/1090
77	1 (**B**ackground)	1 (**B**ackground)	1 (**B**ackground)	1 (**B**ackground)		8.5A2 sterile SN 11+-94
93	1 (3 weak) (**B**)	1 (3 weak) (**B**)	1 (**B**)	1 (3 weak) (**B**)		I 19-AJ 27/12 89 concentrated 22/1/90
108	1, 12	1 (**B**) (3 weak)	1 (**B**)	1,12	1,12	+ II 1066 31.3.
112	1 (3,7,8,9 weak)	1 (3,7,8,9 weak)	1, 8, 9 (**B**)	almost recognize all	1,3,7,8,9,12,13	II - 874 4/5-89
175	2	2	2	2, (1 weak)		16-3D10 13.6.89 3?
48	2, (1 weak)	2	2, (1 weak)	2, (1 weak)	12	15 - 1B9 19/3-90 + N3
53	2, (1 weak)	2, (1 weak)	2, (1 weak)	2, (1 weak)		15 4E3 26/2-90
54	2, (1 weak)	2, (1 weak)	2, (1 weak)	2, (1 weak)		14-4E3 19/3-90 + N3

Notes: the numbers from column 2-5 are the abbreviations of different BG-Fc fusion proteins (for example, 8 for BG8-Fc fusion protein, 9 for BG9-Fc fusion protein etc.).

6.3.5.1 Most BG mAbs recognized BG8-9

More than two thirds of the 63 BG mAbs could recognize both BG8 and BG9, which is reasonable as there is only one amino acid difference (R20C) in the Ig-V domains cloned into BG-Fc fusion protein between BG8 and BG9 (Figure 6.8). Among these BG mAbs, a few of them only recognized BG8-9, while others recognized additional BG proteins. To summarize, all the patterns that included BG8-9 were: BG8-9, BG8-9-12, BG8-9-13, BG8-9-12-13, BG4-8-9-12, BG4-8-9-13, BG4-8-9-12-13, BG1-8-9-12-13, and BG1-4-8-9-12-13.

From the sequence homology, it is easy to understand why some mAbs could recognize BG8, 9, 12 and 13, but it is not clear why some mAbs could recognize BG1 or BG4 together with BG8-9. As both coloured in the amino acid sequence alignments and the structure modeled for BG8 (Figure 6.8), BG8-9-12 are highly conserved with very few amino acid differences labeled blue (R20C, V29A, N31S). Therefore, the loop region, where the two different amino acid residues coloured in blue exist (V29A, N31S), could be the only area where the epitope for BG mAb distinguishing BG8-9 to BG12 might bind to. BG13 is very similar to BG8-9-12 in sequence except two stretches involving nine amino acids differences due to micro-recombination (either by double reciprocal recombination or by gene conversion) between BG8-9-12 genes and BG6 (Chattaway *et al.*, 2016), and another three amino acid differences. The epitopes that these BG mAbs distinguish BG8-9-12 from BG13 might be derived from the two stretches with or without another two amino acids (G49, F50) where the variable amino acids in BG13 were labeled red in figure 6.8. However, it is hard to see any regions based on linear sequences that are conserved in the other patterns (BG4-8-9-12, BG4-8-9-13, BG4-8-9-12-13, BG1-8-9-12-13, BG1-4-8-9-12-13) while different from other BG genes. Crystal structures of the Ig-V domains of BG proteins might provide further information.

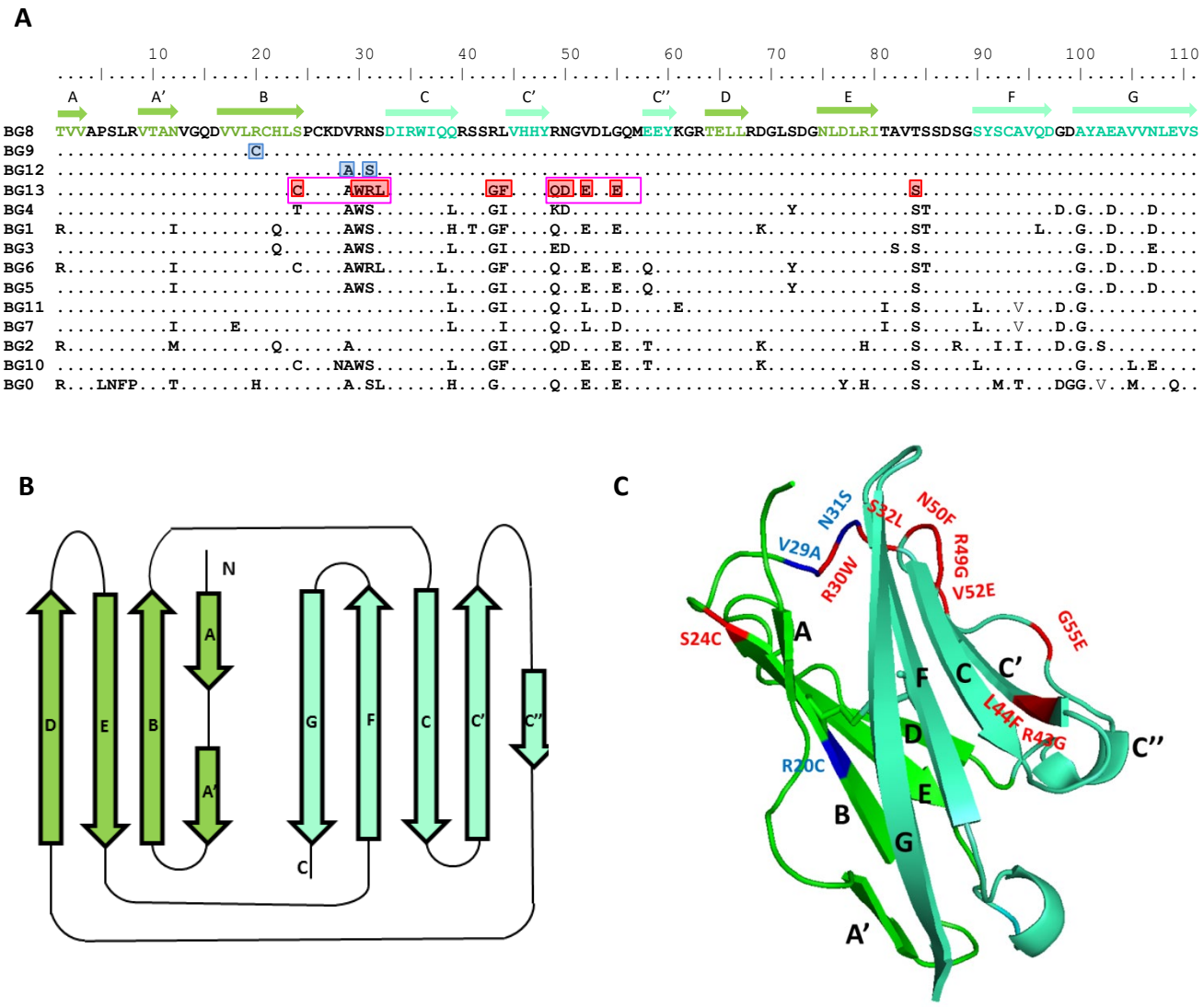


Figure 6.8 Amino acid sequence alignments of the 14 Ig-V domains of BG genes in B12 haplotype, along with structural models of the Ig-V domains with the location of potential epitopes that certain BG mAbs recognizing BG8-9 and related patterns bind to. (A), letters indicate amino acids by single letter code, dots indicate identities with BG8 sequence, residues that differ to BG8 in BG9 and BG12 were labeled in blue, residues that differ between BG8-9-12 and BG13 are labeled in red in BG13 sequence with the two stretches (Chattaway *et al.*, 2016) marked in purple boxes. (B), the β-strands of the V region indicated by arrows in the top panel are coloured dark green for one face of the domain and light green for the other face. (C), the structure built on BG8 with the positions of residues labeled in red show potential epitopes where BG mAbs distinguish BG8-9 from BG12 derived, and in blue distinguish BG8-9-12 from BG13.

6.3.5.2 Three flasks of BG mAb supernatants recognized BG1-6-13

Three flasks of BG mAb supernatants, probably from the same hybridoma clone 'I-18D11', were found to recognize BG1-6-13. As mentioned above, in the Ig-V domain, BG6 and BG13 share two stretches labeled red (position 24-32) and blue (position 49-55), respectively (Figure 6.9), which could be epitopes that certain BG mAbs bind to. Adding BG1 into consideration, based on the amino acid sequence alignments, it seems that only epitopes that relating one amino acid (Q49) from one stretch above together with another two amino acids (G43 and F44) labeled hot pink on BG1 sequence may distinguish BG1-6-13 from others. The three amino acids were labeled on the structure modeled for BG8 with the corresponding colours above.

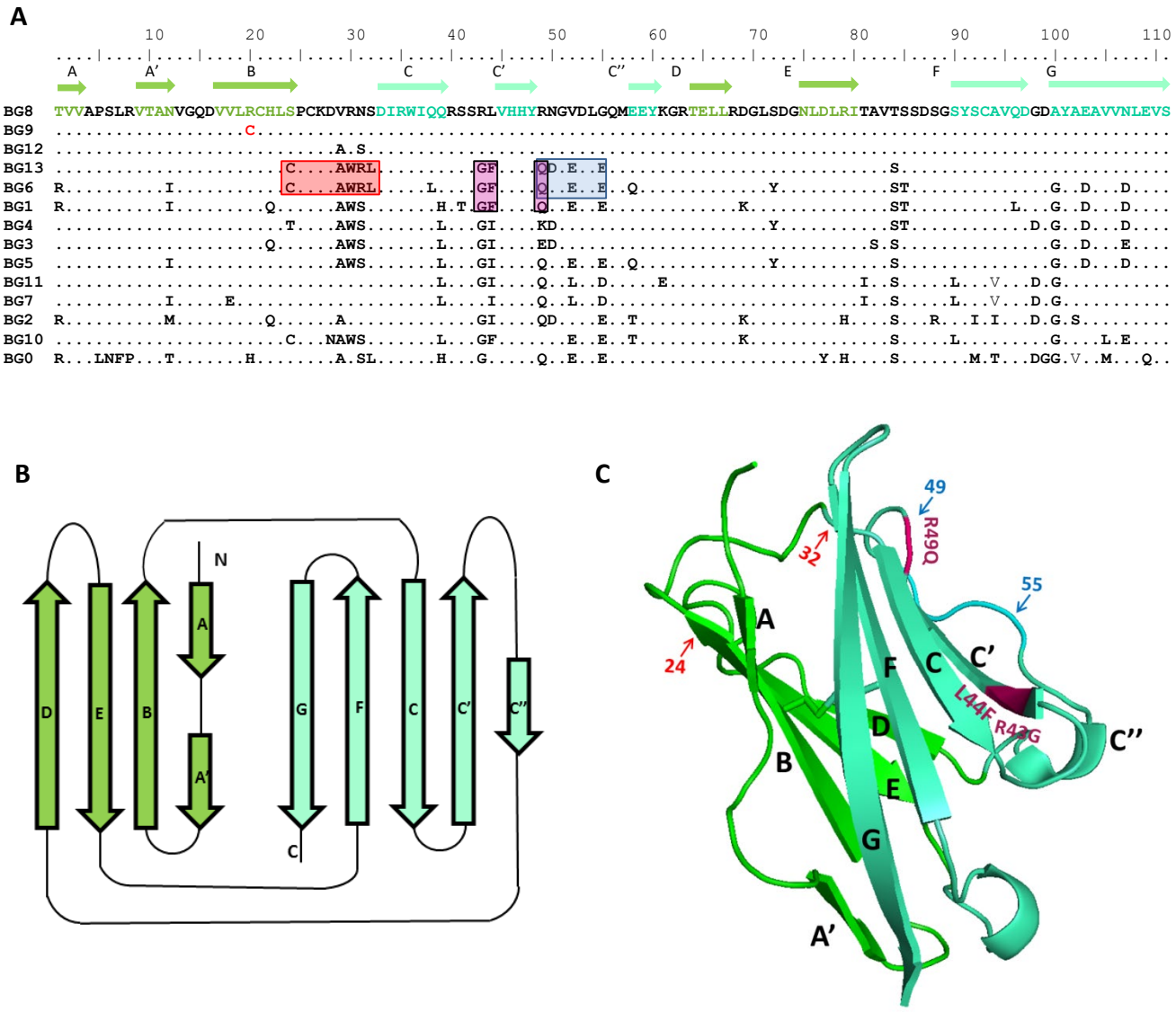


Figure 6.9 Amino acid sequence alignments of the 14 Ig-V domains of BG genes in B12 haplotype, along with structural models of the Ig-V domains with the location of potential epitopes that certain BG mAbs recognizing BG1-6-13 might bind to. (A), letters indicate amino acids by single letter code, dots indicate identities with BG8 sequence, the two stretches shared by BG13 and BG6 are marked in red (from position 24 to 32) and blue (from position 49 to 55) boxes (Chattaway *et al.*, 2016), and the potential epitope that certain BG mAbs recognizing BG1-6-13 might bind to are marked in hotpink box purple (including position 43, 44 and 49). (B), the β-strands of the V region indicated by arrows in the top panel are coloured dark green for one face of the domain and light green for the other face. (C), the structure built on BG8 with the positions of residues labeled accordingly with (A).

6.3.5.3 Four flasks of BG mAbs supernatants recognize BG2-5-7-11

BG7 and BG11 are very similar in sequence. They differ from other BG genes in two large regions labeled blue (position 49-55, 81-94) in amino acid sequence alignments (Figure 6.10). BG7 is one of the major haemopoietic BG genes in blood cells, which has been found highly expressed in B cells of B12 haplotype chicken. Therefore, it would be reasonable to find a mAb which can specifically bind to BG7 or BG7-11. However, we only observed four flasks of BG mAb supernatants (which actually might come from the same hybridoma clone '15-3D7'), and instead of binding to BG7-11, they recognized BG2-5-7-11. According to the sequence alignments, it seems that only the two amino acids on the two ends of C' strand (I44 and Q49) labeled in red can distinguish BG2-5-7-11 from other BG genes. One scenario could be the epitope that these four BG mAbs bind to was the whole C' strand (I44VHHYQ49), and position 44 and 49 are the most critical amino acids determining the binding.

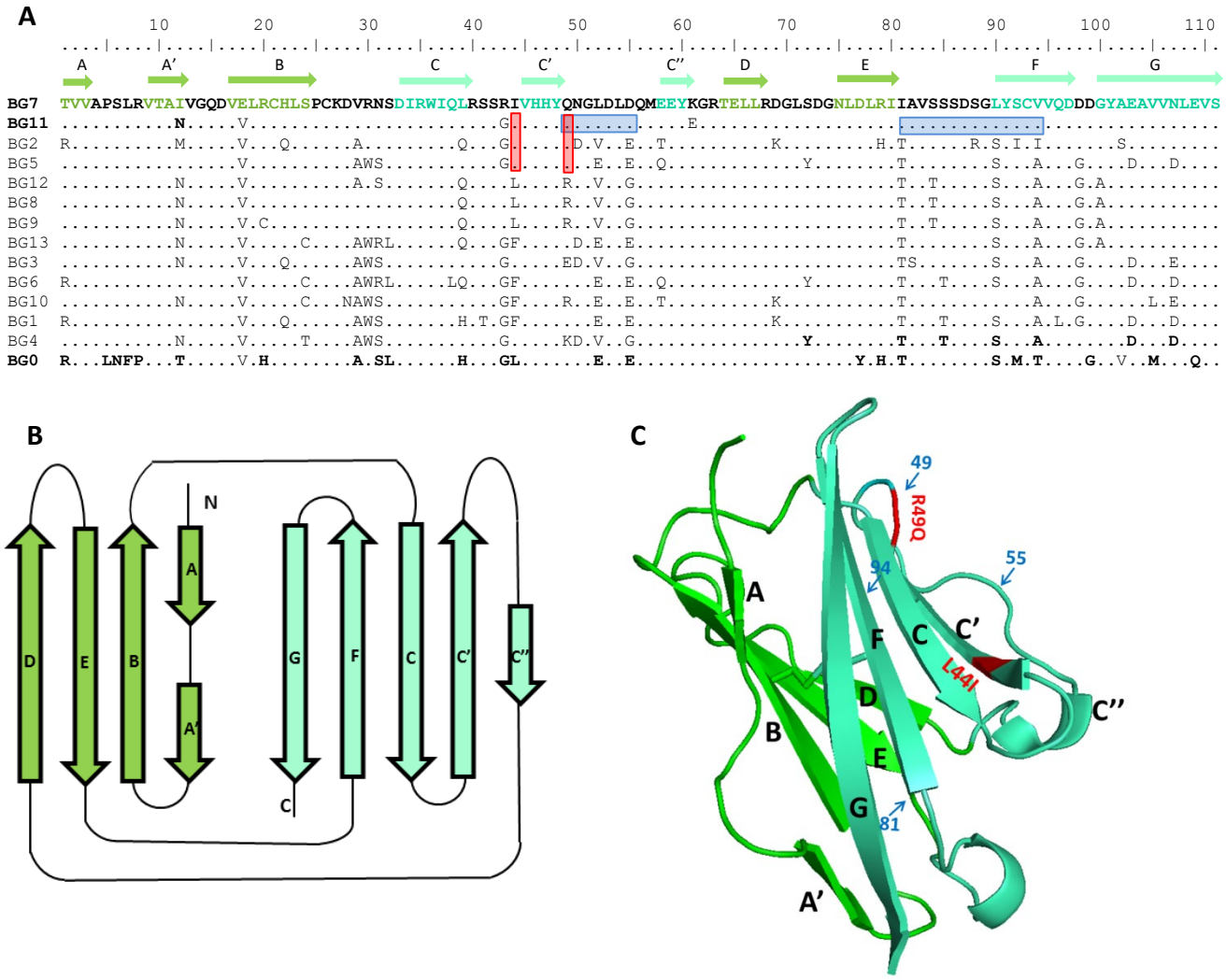


Figure 6.10 Amino acid sequence alignments of the 14 Ig-V domains of BG genes in B12 haplotype, along with structural models of the Ig-V domains with the location of potential epitopes that certain BG mAbs recognizing BG7-11 and BG2-5-7-11 might bind to. (A), letters indicate amino acids by single letter code, dots indicate identities with BG7 sequence, the two regions (from position 49 to 55, position 81 to 94) which distinguish BG7-11 to other BG genes are marked in blue boxes, and the two amino acids (I44 and Q49) on the end of the potential epitope which might distinguish BG2-5-7-11 to other BG genes are marked in red boxes. (B), the β-strands of the V region indicated by arrows in the top panel are coloured dark green for one face of the domain and light green for the other face. (C), the structure built on BG8 with the positions of residues labeled accordingly with (A).

6.3.5.4 Two flasks of BG mAb supernatant recognize BG4-5

Two flasks of BG mAb supernatant, which might come from the same hybridoma clone '14-7C11', recognized both BG4 and BG5 with some background binding to BG2. No clue could be found on the sequence alignment, because no residue was observed being unique for BG4-5 compared to all the other BG genes.

6.3.5.5 A few BG mAbs recognize single BG proteins

There were five flasks of BG mAb supernatants (probably from the same hybridoma clone '14-8B5') specifically recognizing BG5. Four flasks of different hybridomas ('16-3D10', '15-1B9', '15-4E3', '14-4E3') recognize BG2. Another four flasks each from different hybridomas ('8.5A2', 'I-19-AJ', 'II-1066', 'II-874') recognize BG1. However, most of them were observed with certain level of background binding except two flasks that recognized BG2. As shown in figure 6.11, BG2 has seven amino acids that differ to all the other BG genes at particular positions with red letters in sequence alignments, and there are two regions in red box (position 49-58, 88-94) which could comprise the BG2 epitope. For BG1, there are only two amino acids different to all other BG genes at these two positions, but the loop region between strand C and C' labeled in blue box might generate the epitope that some BG mAbs recognizing BG1 only. On the contrary, there is no position in BG5 sequence showing amino acid difference to all other BG genes, therefore, it is hard to find any clue from the linear sequence information about which regions these BG mAbs specifically bind to on BG1.

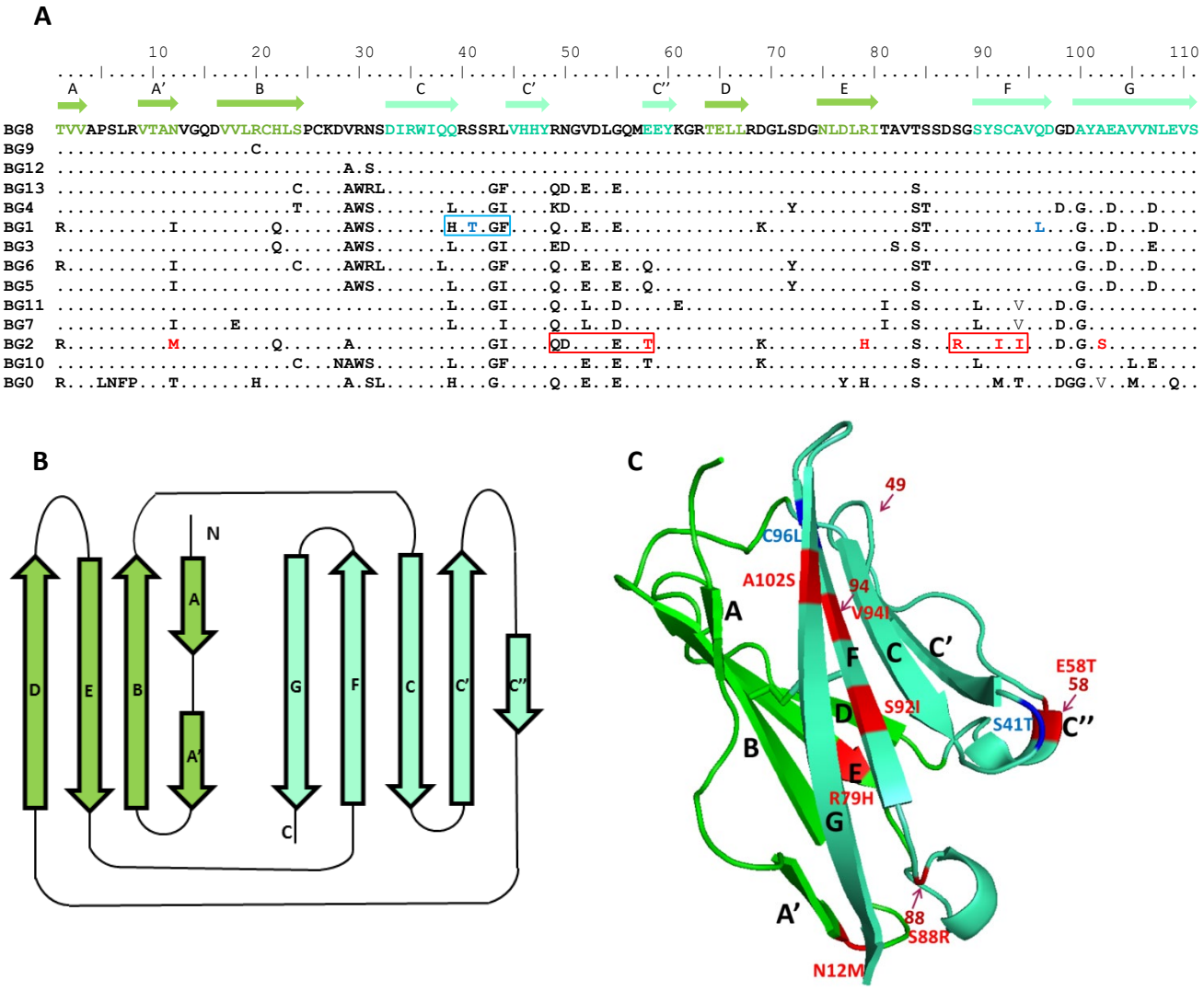


Figure 6.11 Amino acid sequence alignments of the 14 Ig-V domains of BG genes in B12 haplotype from, along with structural models of the Ig-V domains with the location of potential epitopes that certain BG mAbs recognizing BG1 and BG2, respectively, might bind to. (A), letters indicate amino acids by single letter code, dots indicate identities with BG8 sequence, the amino acids unique for BG1 and BG2 are in blue and red colours, respectively, and the potential epitopes that certain BG mAbs specifically recognizing BG1 and BG2 might bind to are marked with blue (loop region between strand C and C': position 39 to 44) and red (two regions: position 49 to 58, position 88 to 94) boxes, respectively. (B), the β-strands of the V region indicated by arrows in the top panel are coloured dark green for one face of the domain and light green for the other face. (C), the structure built on BG8 with the positions of residues labeled accordingly with (A).

6.3.5.5.1 Summary

All 290 flasks of BG mAb supernatants were screened by the first round sandwich ELISA for three times (with all the 14 BG-Fc fusion protein supernatant mixed together as antigens), and 63 flasks were found interacting with BG-Fc fusion proteins. Then the 63 flasks of supernatants were screened by the second round sandwich ELISA (with individual BG-Fc fusion protein supernatant as antigen) at least three times. The results showed interesting patterns, with 65% BG mAb supernatants recognizing BG8-9.

6.3.6 Thirty eight flasks of BG mAb supernatants were picked for SDS-PAGE & Western blot to determine the antibody specificities

In total, 38 flasks of BG mAb supernatants were selected from the 63 BG mAbs characterized by sandwich ELISA, and were further characterized by SDS-PAGE (both reduced and non-reduced gel) and western blot. Among the 63 flasks of BG mAbs, many of them are most likely from the same hybridoma clones as the records on the flasks showed the same names, as well as the same results from sandwich ELISA tests. For example, BG mAb229 and BG mAb158 both were noted with '14 7C11' and they both recognized BG4 and BG5 in sandwich ELISA. Five BG mAb supernatants, BG mAb195, mAb199, mAb200, mAb201 and mAb236, were noted with 'I 2C10' and recognized BG8 and BG9 with weak interaction with BG12 and BG13. To reduce the unnecessary repeat of western blot assays and avoid the wastes of precious BG mAb supernatants, only 38 flasks of BG mAbs supernatants were picked and used in SDS-PAGE and western blot (Table 6.4). For instance, two BG mAbs, mAb158 and mAb201, were picked to represent the other mAbs from the examples above for SDS-PAGE and western blot study.

Both reduced and non-reduced SDS-PAGE gels were performed at least three times for each sample, and the final results were summarized by comparing with sandwich ELISA results in the following section (Table 6.5) with WB results of two samples displayed in figure 6.12.

Table 6.4 A total of 38 BG mAbs (labelled yellow in the first column) were picked for SDS-PAGE and western blot test

BG mAb ID	ELISA results	Notes on the flask	Volume	Concentrated	Quality
229	4,5, (2 weak)	14 7C11 SN+N3 28.4.89	28 mL		
158	4,5 (2,3 weak)	14-7C11 pool Sept. 91. +N3	155 mL		
110	8,9	2C-10-2 7/2	100 mL		
137	8,9	II 301 31/3 concentrated ???pool 22/7 91 +N3	55 mL	concentrated	
148	8,9	II 349 14/4 89 concentrated + ????? 18/7 91 added	50 mL		
153	8,9	Fu II 349 4/5.89 +0.5%N3	50 mL		
195	8,9,(12,13 weak)	I 2C.1D 28/12 87 1/11.90 concentrated	25 mL	concentrated	
199	8,9,(12,13 weak)	I-2C10-1 7.2.	110 mL		
200	8,9,(12,13 weak)	I2C10 20.11. SN + N3	125 mL		
201	8,9,(12,13 weak)	I 2C10 30/10 concentrated 19/11.90	60 mL	concentrated	
236	8,9,(12,13 weak)	I 2C10 +N3 9/4 91	60 mL		
128	1,6,13	I 18-D.11 (28/2.90) concentrated 30/10.90	40 mL	concentrated	
129	1,6,13	I 18D11 30/10 concentrated 16/11.90	25 mL	concentrated	
130	1,6,13	I 18D11 20/11. concentrated pool 20/11 90	50 mL	concentrated	
8	8,9,12	II-427 + N3 10/10 90	15 mL	No	good
15	8,9,12	II-240 + N3 10/10 90	0.04 mL	likely	good
55	8,9,13	I8D8-3 15.5 pool	25 mL		
131	8,9,13	I 8D8-33 28/12 89 concentrated pool 6/11 90	50 mL	concentrated	
57	8,9,13	I8D8-3 pool 1 Feb 90	17 mL		
75	8,9,13	J8D8 A 8.5.89 (filtered 24/10/90 due to ?????? G	15 mL		
56	8,9,13 (1,12 weak)	I8D8-3 pool 1 Feb 90	20 mL		
100	8,9,13	I -2C10 2/4.91	430 mL		
127	8,9,13	I - 18C4-3 7.2.	120 mL		
245	8,9,13	I-18C4-4 7.2.	100 mL		
193	8,9,13	I-1A8-2 7.2.	17 mL		
235	8,9,13 (**B**)	I-1A8-1 7.2.	50 mL		
194	8,9,13	I-2E3 7.2.	30 mL		
237	8,9,13	I-2E3-2 7.2.	60 mL		
189	8,9,13 (**B**)	? F16g 9. (18-6G2) +N3 18/6.90	125 mL		
196	8,9,13 (**B**)	mAbg 9 +0.1%N3 (18-6G2) 25/6-90	55 mL		
9	8,9,12,13	II-409 + N3 10/10 90	0.5 mL	likely	good
286	8,9,12,13	mAbg 9 (18-6G2)			
289	8,9,12,13	I-8D8 mAbg1 9/4			
91	8,9,12,13(**B**)	I-18H6 9.2.	15 mL		
123	1,8,9,12,13	I 17A8-1 16/689 concentrated 20/11.90	2.5 mL	concentrated	
151	1,8,9,12,13	FII8 II 431 25.5.85	75 mL		
251	1,8,9,12,13	I-18H6-2 7.2.	100 mL		
47	2,5,7,11	15 - 3D7 pool 1 Lab 80 95	3.5 mL	No	good
49	2,5,7,11	15 - 3D7 pool 1 Feb 80	9 mL	No	good
176	2,5,7,11,(1 weak)	15-3D7 SN +0.1%N3 13/3.90	50 mL		
287	2,5,7,11?	15-3D7 pool 2. Feb			
190	4,8,9,13	18-6G2 1/6-90 +0.1%N3	170 mL		
202	4,8,9,13	18-6G2 pool. May 90 +0.1%N3	200 mL		
242	4,8,9,13	I-17B8 21/4.89	4 mL		
243	4,8,9,13	I 17B8-11 SN+N3 30.6.83	50 mL		
10	4,8,9,12	II-390 + N3 8/10 90	13 mL	No	good
37	4,8,9,12,13	I 17B8-2 27.5 Pool	30 mL	No	good
38	4,8,9,12,13	I 17B8-1 19.5 Pool	22 mL	No	good
39	4,8,9,12,13	(I) 17B8 12.5.89	27 mL	No	good
231	1,4,8,9,12,13	I8IE10 SN+N3 28.6.89	80 mL		
107	5	14-8B5 3/1-89	95 mL		
154	5	14-8B5 3/1-89	120 mL		
157	5	148B-5 30/10 concentrated 18/11.90	10 mL	concentrated	
155	5 (**B**)	148B5 20.11 SN+N3	100 mL		
156	5 (**B**)	14-8B5 28/12.89 concentrated 31/1090	45 mL	concentrated	
77	1 (**B**ackground)	8.5A2 sterile SN 11+-94	1 mL	concentrated	contaminated
93	1 (3 weak) (**B**)	I 19-AJ 27/12 89 concentrated 22/1/90	85 mL	concentrated	
108	1, 12	+ II 1066 31.3.	3 mL		contaminated
112	1 (3,7,8,9 weak)	II - 874 4/5-89	75 mL		
175	2	16-3D10 13.6.89 3?	18 mL		
48	2, (1 weak)	15 - 1B9 19/3-90 + N3	9 mL	No	good
53	2, (1 weak)	15 4E3 26/2-90	33 mL	No	good
54	2, (1 weak)	14-4E3 19/3-90 + N3	13 mL	No	good

6.3.7 Comparison of the ELISA and WB results

The comparison between sandwich ELISA results and western blot results showed quite complicated situations with data presented in table (Table 6.5).

6.3.7.1 A few mAbs show same results in ELISA and both reduced and non-reduced WB

Only three mAbs show the same results in ELISA and both reduced and non-reduced WB, indicating that they recognize linear epitopes. As shown in table 6.5, BG mAb137 (from hybridoma 'II 301') specifically recognize BG8 and BG9. Two hybridomas, BG mAb175 ('16-3D10') and mAb53 ('15 4E3'), both recognize BG2.

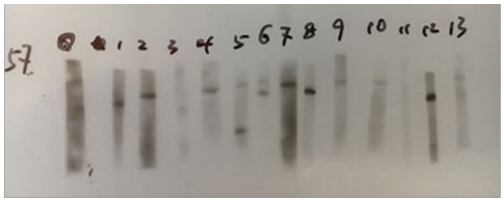
6.3.7.2 Some mAbs show same results in ELISA and non-reduced WB but different results in reduced WB

Eleven BG mAbs show same results in ELISA and non-reduced WB but different results in reduced WB, indicating that they may recognize conformational epitopes (an example of BG mAb48 in figure 6.12). Four flasks of BG mAb supernatants, mAb154, mAb155, mAb156 and mAb158, which probably came from the same hybridoma '14-8B5', all recognized BG5 in ELISA and WB non-reduced gel. Three flasks from three different hybridomas, mAb48 ('15-1B9'), 53 ('15 4E3') and 54 ('14-4E3'), all recognize BG 2 in ELISA and WB non-reduced gel. Another four flasks from four different hybridomas recognize different BG proteins in ELISA and WB non-reduced gel, which were: mAb8 ('II-427') recognizing BG8-9-12, mAb37 ('I 17B8') recognizing BG4-8-9-12-13, mAb49 ('15-3D7') recognizing BG2-5-7-11, and mAb158 ('14-7C11') recognizing BG4-5.

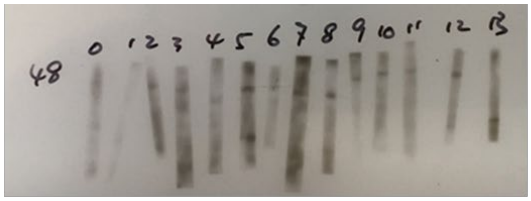
6.3.7.3 A lot of mAbs show different results in ELISA and WB

More than half of the BG mAbs tested for WB showed different results compared to the ELISA (an example of BG mAb57 in figure 6.12). Some mAbs have complicated patterns. For example, many mAbs recognizing BG8-9-13 in ELISA, interact with BG8-9-12 in non-reduced WB, and show strong band for BG12 in reduced WB and weak bands for almost all the other BG proteins. There are also some mAbs that recognize certain numbers of BG proteins in ELISA, but fewer BG proteins in non-reduced WB, and different results for reduced WB. For example, mAb251 ('I-18H6-2') recognize BG1-8-9-12-13 in ELISA, but only binds to BG8-9-12 in non-reduced WB, and binds to all BG proteins except BG1 and

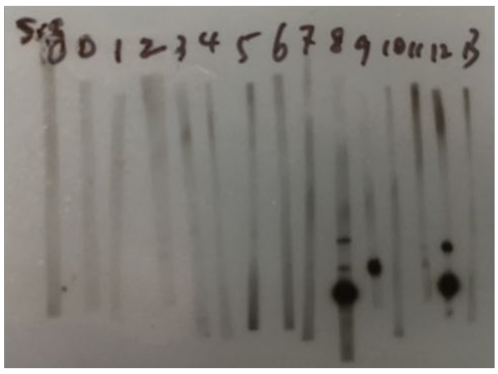
BG3 in reduced WB. One possibility of explaining this observation is the affinity of these mAbs binding to corresponding BG proteins differ in ELISA and non-reduced WB. Another possibility is that these mAb supernatants might not be from a single clone. If the supernatants contained two or more different mAbs, which all recognize certain BG proteins in different ways, it may turn out different results when using different methods.



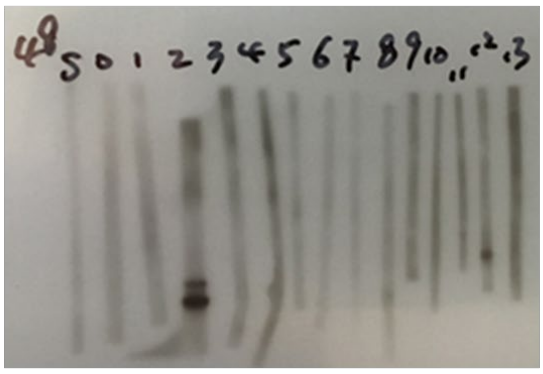
BG mAb57
Reduced condition, exposure 10''
Strip order: BG0-Fc, 1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13



BG mAb48
Reduced condition, exposure 10''
Strip order: BG0-Fc, 1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13



BG mAb57
Non-reduced condition, exposure 10''
Strip order: B7H6-Fc, BG0-Fc, 1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13



BG mAb48
Non-reduced condition, exposure 10''
Strip order: B7H6-Fc,BG0-Fc, 1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13

Figure 6.12 Examples of two BG mAb supernatants showing different results in western blot reduced and non-reduced conditions. The upper samples (the number above each strip with 0 for BG0-Fc fusion protein, 1 for BG1-Fc fusion protein, and so on) were incubated in SDS-containing buffer at 85^0C for 5 min with DTT ('boiled and reduced'); the bottom samples (the number above each strip with Sig/S for B7H6-Fc fusion protein, 0 for BG0-Fc fusion protein, 1 for BG1-Fc fusion protein, and so on) were boiled but without DTT in buffer ('boiled and non-reduced'). The strips were incubated with BG mAb57 (the left two pictures) and BG mAb48 (the right), washed, and incubated with secondary antibody (goat anti-mouse IgG-HRP), followed by X-ray film exposure for 10 seconds. The results showed that the two BG mAbs recognized BG-Fc fusion proteins differently under reduced and non-reduced conditions. For example, BG mAb57 recognized BG8-Fc, BG9-Fc and BG12-Fc with weak reaction with BG11-Fc under non-reduced condition, but almost recognized every BG-Fc fusion proteins under reduced condition.

Table 6.5 Comparison of ELISA and western blot (both reduced and non-reduced gels) characterizing a total of 38 BG mAbs

mAb ID	ELISA result	WB non-reduced	WB reduced	Notes on the flask
158	4,5 (2,3 weak)	4,5	W all E 7	14-7C11 pool Sept. 91. +N3
110	8,9	8,9,12	8,9,10,12,13	2C-10-2 7/2
137	8,9	8,9	8,9	II 301 31/3 concentrated ???pool 22/7 91 +N3
201	8,9,(12,13 weak)	8,9,12	8,12/W all	I 2C10 30/10 concentrated 19/11.90
130	1,6,13	W1,2,3,5,7,8,9	all E 1, 3	I 18D11 20/11. concentrated pool 20/11 90
8	8,9,12	8,9,12	8,11,12/W all E 1,3	II-427 + N3 10/10 90
131	8,9,13	8,9,12/W11	12/W all E 1,3	I 8D8-33 28/12 89 concentrated pool 6/11 90
57	8,9,13	8,9,12/W11	12/W all E 1,3	I8D8-3 pool 1 Feb 90
56	8,9,13 (1,12 weak)	8,9,12	12/W all E 1,3	I8D8-3 pool 1 Feb 90
100	8,9,13	8,9,12	12/W all E 1,3	I -2C10 2/4.91
245	8,9,13	8,9,12	8,12/W all E 1,3,**11**	I-18C4-4 7.2.
75	8,9,13	8,12/W9	W2,8,12/B	J8D8 A 8.5.89 (filtered 24/10/90 due to ?????? Growth by Fiona
55	8,9,13	8,12	W0,2,7,8,9,10,12,13	I8D8-3 15.5 pool
237	8,9,13	8,9,12	8,12/W 2,6,8,10,13	I-2E3-2 7.2.
193	8,9,13	8,9,11,12	12/W all E 1,3	I-1A8-2 7.2.
235	8,9,13 (**B**)	8,12/W9,11	12/W all E 1,3	I-1A8-1 7.2.
189	8,9,13 (**B**)	8,12/W4,9	8,12/W all E 1,3	? F16g 9. (18-6G2) +N3 18/6.90
289	8,9,12,13	8,9,12	all E 1, 3	I-8D8 mAbg1 9/4
91	8,9,12,13(**B**)	8,9,11,12	12/W all E 1,3	I-18H6 9.2.
151	1,8,9,12,13	8,9,12/W10	8,12/W 9,13; WW 4,5,6	FII8 II 431 25.5.85
251	1,8,9,12,13	8,12/W9	12/W all E 1,3	I-18H6-2 7.2.
49	2,5,7,11	5,7/W2,11	5,11/W all E 1,3	15 - 3D7 pool 1 Feb 80
176	2,5,7,11,(1 weak)	1,5,7,11	2,4,5,11,12,13/W all E 1,3	15-3D7 SN +0.1%N3 13/3.90
202	4,8,9,13	4,8,9,12/W13	8,12/W all E 1,3	18-6G2 pool. May 90 +0.1%N3
243	4,8,9,13	4,8,9,12/W11	8,9,12/W all E 1,3	I 17B8-11 SN+N3 30.6.83
10	4,8,9,12	8,9,12/W11	2,8,12/W3,4,5,6,7,9,10	II-390 + N3 8/10 90
37	4,8,9,12,13	4,8,9,12,13	8,12/W4	I 17B8-2 27.5 Pool
231	1,4,8,9,12,13	4,8,9,12	8/W2,4,5,6,12	I8IE10 SN+N3 28.6.89
107	5	4,5,8,9,12	2,5,8/W all **E 1**	14-8B5 3/1-89
154	5	5	2,4,5,6,7,9,10,12/W3,8,11	14-8B5 3/1-89
157	5	5/B	All/W 1	148B-5 30/10 concentrated 18/11.90
155	5 (**B**)	5/W2,4,7,8,11,12	All	148B5 20.11 SN+N3
156	5 (**B**)	5/W4,6,7,8,11,12	All	14-8B5 28/12.89 concentrated 31/1090
93	1 (3 weak) (**B**)	8,9,11,12	W2,4,5,6,7,8,9,12,13	I 19-AJ 27/12 89 concentrated 22/1/90
108	1, 12	nothing	**nothing**	+ II 1066 31.3.
112	1 (3,7,8,9 weak)	8,9,11,12	7,8,12/W2,4,5,6,9,10,13	II - 874 4/5-89
175	2	2,5,8	2	16-3D10 13.6.89 3?
48	2, (1 weak)	2/W12	all	15 - 1B9 19/3-90 + N3
53	2, (1 weak)	2	2/B	15 4E3 26/2-90
54	2, (1 weak)	2	2/S6,7,12/W0,4,5,8,9,10,11	14-4E3 19/3-90 + N3

Notes: Same colours labeled indicate same pattern. ‘B’ for strong background reading, ‘W’ for weak reaction, ‘E’ for except (e.g. ‘8,11,12/W all E 1, 3’ means ‘strong band for BG8, 11 and 12, weak bands for all the other BG proteins except BG1 and BG3’).

6.4 Discussion

In order to confirm many features found about BG genes in previous studies at protein level, as well as to further characterize BG proteins, for instance, tissue distribution, ligand interaction, crystal structures etc., two major studies were carried out in this chapter.

Firstly, the Ig-V domains of all 14 BG genes from B12 haplotypes were cloned into SigpIg Plus vectors. The SigpIg Plus vector itself has a CD33 signal sequence, an MCS, a linker peptide and the human IgG1 Fc domain. Once the targeted genes were successfully inserted into the vector, SigpIg Plus plasmids could be transfected into eukaryotic cell lines to produce Fc fusion proteins. Fourteen BG-Fc SigpIg Plus plasmids were transfected into HEK293 cells separately and the cells were selected by G418 to make stable cell lines which yielded high quantity of BG-Fc fusion proteins. The process of making stable cell lines were time consuming and with certain uncertainties. For some cell lines, for example BG0-Fc and BG6-Fc, it took about four months to accomplish. Even then the fusion proteins were not produced at high yield. However, this might be expected, because our stable cell lines were not from single clone nor selected for high yield. Another possibility was that each BG-Fc SigpIg Plus plasmid has different sequence due to the BG domain, which might influence the plasmid integration into HEK293 cell genome. Therefore, if large quantity of BG-Fc proteins are needed in the future, single clones that produce high yields should be selected for stable cell lines production.

Secondly, a sandwich ELISA was developed to characterize 290 flasks of BG mAb supernatants using BG-Fc fusion proteins. In order to examine such large number of BG mAbs against 14 different BG-Fc fusion proteins, two rounds of sandwich ELISAs were established. During the first round, all 14 BG-Fc fusion proteins were mixed to check if they could be recognized by any of the 290 BG mAbs. Only those BG mAb showing positive results were selected into the second round of sandwich ELISA where individual BG-Fc fusion proteins were used. All the tests were repeated at least twice to ensure the sandwich ELISA results, and one repeat was done using a commercial kit, ‘goat anti-mouse IgG microplate’ (R&D). We have known that at least two BG mAbs (mAb288 and mAb290) were IgM (Figure 6.7) (Salomonsen *et al.*, 1991), so these BG mAbs showing positive using our homemade kit with coated antibody ‘goat anti-mouse IgG, IgA, IgM (H+L)’ (BioFX) might be negative using commercial kit. However, the results showed no difference, although our homemade kits had a higher background reading. These two IgM mAbs were both negative

using either the homemade or the commercial kit. One explanation would be that our homemade kit couldn't detect IgM; second possibility is that the IgM were not stable during the 30 years of storage at 4°C second; a third possibility is that the two IgM mAbs previously found recognizing BG proteins on peripheral blood lymphocytes of line CB (B12) chicken using flow cytometry bind to some special epitope determinants that our BG-Fc fusion proteins doesn't have. For example, our BG-Fc fusion proteins are all homodimers, while in chicken blood, BG proteins may be heterodimers on the cell surface.

In total, 63 flasks of BG mAb out of 290 were confirmed to recognize BG-Fc fusion proteins with some interesting patterns. It is not surprising that more than two thirds of the mAbs didn't recognize BG-Fc fusion proteins, and at least two possibilities could contribute to the fact. First, these BG mAbs were made using erythrocytes; many mAbs might recognize particular epitopes from BG proteins that are only present on the cell surface (homodimer, heterodimer, or other unknown structures), while we have used secreted homodimer BG-Fc fusion proteins which may only partially represent the real structure. Second, these BG mAbs were made against erythrocytes and lymphocytes of B19 or B21 haplotypes according to Prof. Salomonsen (personal communication). BG genes are highly polymorphic and no identical BG gene was found from T cells and B cells between the B12 haplotype and the B19 or B21 haplotypes, so it is very likely that many mAbs can't recognize BG proteins made from B12 haplotype.

Also, because it is unclear which chicken haplotypes were the immunogens, it is not easy to understand all the patterns generated from the sandwich ELISA. Some patterns are interpretable; for example, 65% of the ELISA positive BG mAb supernatants recognized BG8-9. According to the results in chapter four, the dominantly expressed BG genes from peripheral T cells of different chicken lines all belong to the BG8-9-12 clade except line 6_1 (B2), indicating the similarities of these sequences (Chen *et al.*, 2018). Since these BG mAbs were made against erythrocytes, it is reasonable that most mAbs recognize BG8-9. Another BG gene, BG7, was found dominantly expressed in peripheral B cells from line 6_1 (B2) (chapter five) and highly expressed in line CB (B12) (Salomonsen *et al.*, 2014), but only four flasks of BG mAbs were found that recognized BG7. However, there are some patterns that we could not understand at all. For instance, some supernatants recognize BG 4-5, but there is no residue that is common only in BG4 and BG5 according to the amino acid sequence alignments of all 14 BG genes. Also, both the amino acid sequence alignments and the

modeling structure do not provide a simple explanation for these supernatants recognizing BG2-5-7-11 or BG1-6-13. Such mysteries can be only understood with further characterization of these mAbs, or with the real crystal structures of Ig-V domains of BG proteins. Alternatively, since 14 BG-Fc stable cell lines have been made, specific BG mAbs could be made using these BG-Fc fusion proteins if further characterization of the BG mAbs above fails.

To sum up, the work done in this chapter has provided powerful tools for the coming studies to explore BG functions. These BG-Fc fusion proteins can be used in functional assays to check if BG proteins could stimulate γδT cell proliferation. BG-Fc fusion proteins can also be used in staining lymphocytes to see if potential ligands can be found using flow cytometry. BG monoclonal antibodies can be used in tissue distribution study, as well as in proteomics to investigate which molecules that the cytoplasmic tails of BG proteins interact with.

Chapter 7

Conclusion and future work

7.1 The discoveries in this project

Due to the high polymorphism, copy number variation and very few reagents available, research on BG genes and their functions has progressed only slowly. In this project, firstly, a universal primer pair named HU amplifying haemopoietic BG genes was established and applied to systemically study BG gene expression in T and B cells from four different chicken lines (line N, line P2a, line 15I and line 6_1), from which 23 BG genes were found. Secondly, all the dominantly expressed BG genes (the 'functional alleles') in T cells and B cells were compared and analyzed, and only the cytoplasmic tail region is clearly seen to be under selection. Thirdly, essential reagents were developed for further function studies. The most important discoveries and possible interpretations are listed as following.

1. Depending on the haplotype, the number of BG genes found in B cells is not always more than that in T cells. Previously, Salomonsen *et al*., 2014 found only two BG genes expressed in T cells versus five in B cells from line CB chicken (B12 haplotype) (Salomonsen *et al.*, 2014), and the same situation was expected in other chicken lines. However, from all four chicken lines tested in this project, only two chicken lines, line 6_1 and line 15I, showed more BG genes in B cells compared to T cells. The exact same three BG genes were found both in T and B cells from line P2a; and fewer BG genes were found in B cells than in T cells from line N. Also, in previous work, it was not clear about the full-length cDNA sequences, as the primers used were only designed to target partial signal sequence, full Ig-V domain and partial transmembrane region. In this current project all the BG cDNA sequences are nearly full-length, amplified by newly developed primer pair HU. Through comparison of these transcripts, two important features of BG genes were revealed: alternative splicing is observed for most BG cDNA sequences and the exon 2 encoding Ig-V domain can help distinguish different BG genes.

2. The alternative splicing isoforms of BG cDNA sequences, potentially encoding soluble BG proteins, were only observed in B cells. As summarized in figure 7.1, many BG genes detected both in T and B cells from each chicken line are of various alternative splicing isoforms, which are mostly resulted from intron retention where early stop codons are introduced. When the intron is retained in the cytoplasmic tail region, truncated cytoplasmic tails are generated, and when the intron between exon 2 (Ig-V domain) and exon 3

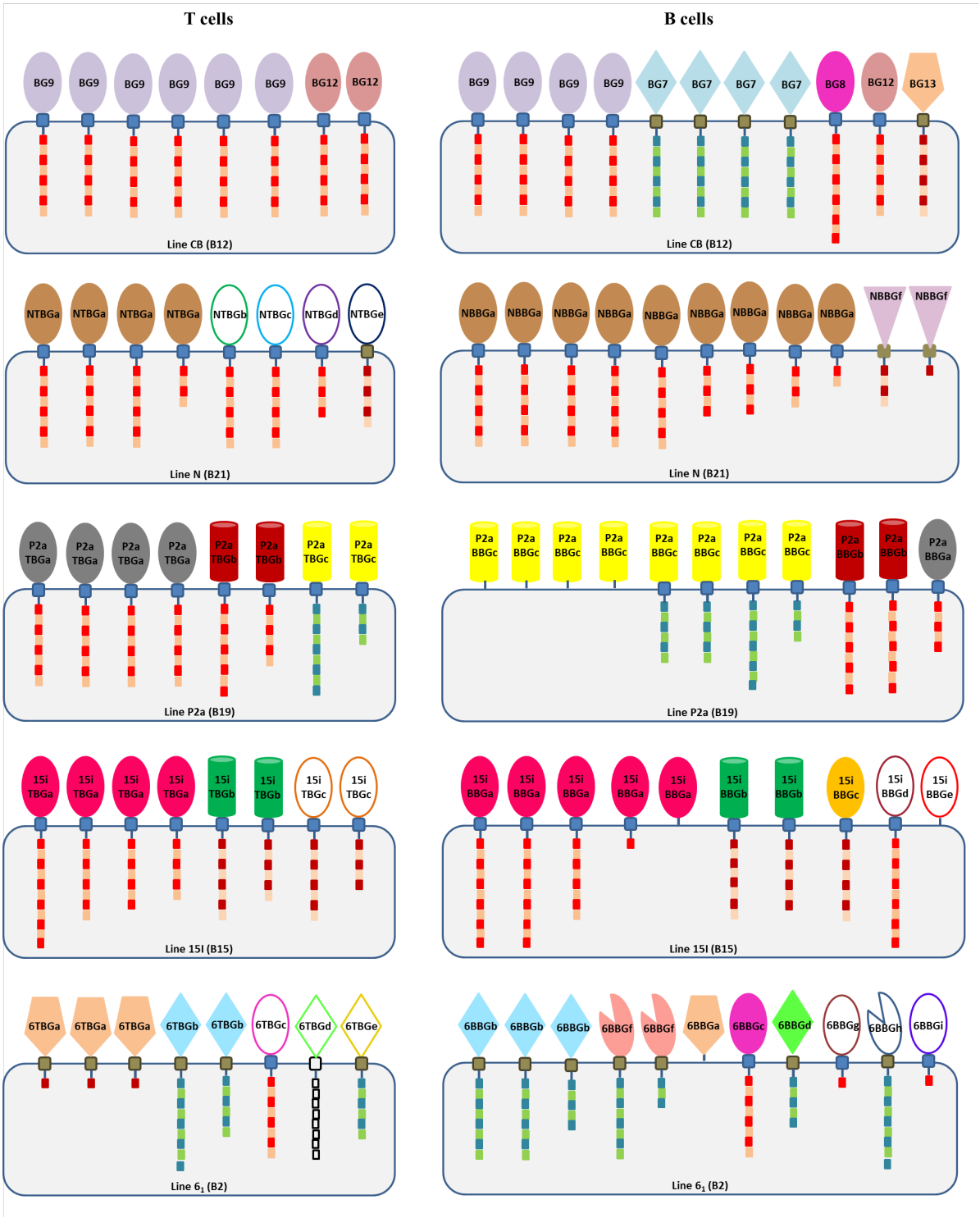


Figure 7.1 Cartoon summary of all BG genes found in T and B cells of five chicken lines.

T cells on the left and B cells on the right with each chicken line specified on the bottom. The same shapes of excellular domain stand for their Ig-V domains belongs to the same group in the phylogenetic tree built on Ig-V domain nucleotide sequences, and if in the same colour, they are the same BG genes. Two types of transmembrane regions are indicated by blue and grey except 6TBGd in T cells which lack the sequence information for transmembrane regions. Two types of cytoplasmic tails are indicated by green/blue (type 1) and red/carnation (type 2) with type 2 having two sub-types, bright red (type 2a) and dark red (type 2b); and the lengths of cytoplasmic tail represent the alternative splicing.

(transmembrane region) is retained, a soluble BG protein might be produced. Interestingly, the cDNA sequences potentially encoding soluble BG proteins were only observed in B cells, indicating important functional differences of the same BG genes in T cells and B cells. Soluble isoforms of many known costimulatory molecules as well as immunoregulatory receptors [e.g. CTLA-4, T cell Ig mucin 3 (Tim-3), CD86, Fas etc.] are more recently understood to be very important in regulating immune response and be associated with autoimmune diseases (Dev *et al.*, 1994; Jeannin *et al.*, 2000; Clayton *et al.*, 2015; Oaks *et al.*, 2018). BG genes are considered to have similar functions as their homologous human BTN molecules which have been shown of great importance in immune regulations. Therefore, it would be useful to bear in mind that there are possibly natural forms of soluble BG proteins secreted from B cells.

3. The true *a* and *d* positions in a heptad repeat of cytoplasmic tails of BG genes forming the α-helical coiled coil are encoded by the fourth and seventh codons of the 21 nucleotide exons. BG proteins are well known as dimers with two cytoplasmic tails in α-helix forming a coiled coil, and we thought that the first and forth codons in the 21 nucleotide exon of BG cytoplasmic tail encode the *a* and *d* positions in the α-helical coil. Helical wheel structures of the cytoplasmic tails of all BG genes analyzed in this project clearly showed a pattern that the forth and seven codons but not the first and forth codons in the 21 nucleotide exon of BG cytoplasmic tail form a hydrophobic interface. There are many amino acids from the first and third codon charged, potentially allowing salt bridges formed between oppositely charged amino acids of the two chains (Aronsson *et al.*, 2015), which exactly satisfied the amino acid requirements for *e* and *g* positions. Understanding the true structures of cytoplasmic tails of BG genes, many potential motifs were displayed with some being tested important in functions of the homologous BTN molecules; for example ‘KKXX’, ‘EWK’ etc., are associated with ER retention (Vantourout *et al.*, 2018).

4. BG genes look like hybrid genes. The phylogenetic trees built on different regions show that some BG genes are clustered into the same group in some regions but divided into different groups in other regions (Chattaway, 2013; Salomonsen *et al.*, 2014), which contributes greatly to the polymorphism of BG genes. In this project, there were 23 different BG genes found from the four chicken lines, and none of them was identical to another. However, phylogenetic trees of each region of the newly discovered BG genes found in this project showed that all the regions clustered into groups that have already been described in

the phylogenetic trees built of 14 BG genes from B12 haplotype (Salomonsen *et al.*, 2014), with the exception in exon 2 (encoding the Ig-V domain), for which new clades were found. Only looking at the three potential functional domains (Ig-V domain, transmembrane region and cytoplasmic tail region) in figure 7.1, it is not difficult to understand the polymorphisms of these BG genes. There are six different types of Ig-V domains demonstrated with different shapes, two types of transmembrane regions with different colours (grey and blue), and two types of cytoplasmic tails (blue and green exons versus red and carnation exons) with the red-carnation one including two sub-types coloured in dark red or bright red.

5. The BG haplotypes of different chicken B haplotypes examined so far are different. The chicken B haplotype should include three parts, the MHC haplotype (BF and BL), BG haplotype and TRIM haplotype. Previously all the standard chicken lines were typed for MHC haplotypes and BG haplotypes using mAbs and antisera (Briles *et al.*, 1982), from which some different B haplotypes were considered to have different MHC haplotypes but the same or similar BG haplotypes. For example, B2 and B12 were considered to have the same BG haplotype by hemagglutination (Simonsen *et al.*, 1982). In this project, although only four chicken lines were sequenced for BG cDNA sequences from T and B cells, all the results show that BG haplotypes are correlated with B haplotypes, and the BG sequences are conserved between different chickens within the same B haplotype. For example, eight BG genes were found in line 6_1 (B2), of which only six are identical to these found previously in line CB (B12). This means that line CB (B12) and line 6_1 (B2) actually have different BG haplotypes. Also it is worth mentioning that it is not clear whether the BG genes in line CB have the same alternative splicing isoforms as these correlated BG genes in line 6_1. More strikingly, 6TBBGa, which is identical to BG13 in line CB, was dominantly expressed in T cells with all the cDNA sequences having intron retention after the first exon of the cytoplasmic tail, resulting in very short truncated cytoplasmic tail. However, in line CB, the dominantly expressed BG gene in T cell is BG9 but not BG13.

6. By comparing all the dominantly expressed BG genes (the 'functional alleles') in T cells and B cells, only the cytoplasmic tail region is clearly seen to be under selection, based on the overwhelming preponderance of non-synonymous changes. The dominantly expressed BG genes in T cells belong to BG8-9-12 clade except the one from line 6_1. The dominantly expressed BG genes in B cells have either type 1 (blue-green tail in figure 7.1) or type 2a (bright red-carnation) cytoplasmic tail. As shown in figure 7.1, apart from line 6_1, all the

dominantly expressed BG genes from the other three chicken lines together with line CB have the same type of Ig-V domain (in shape of ovals), transmembrane region (in blue colour) and cytoplasmic tail (in bright red-carnation). As mentioned already, the dominantly expressed BG gene in line 6_1, 6TBBGa, is identical to BG13 in line CB, though it has different Ig-V domain, which is believed to be a recombination outcome of BG8-9-12 and BG6 (Chattaway *et al.*, 2016). The transmembrane region of 6TBBGa is a different type; and its cytoplasmic tail although belongs to type 2 but in another sub-type. It is not clear why line 6_1 is exceptional to other lines, but the 'functional alleles' of BG genes found in T cells are BG9, NTBGa, PTBGa, 15iTBGa and 6TBGa.

The dominantly expressed BG genes in B cells look more complicated in the Ig-V domains as shown by different shapes (ovals, diamonds and cylinders) in figure 7.1. Both two types of transmembrane regions and cytoplasmic tails were observed. The only message that can be concluded so far is that in some chicken lines, either type 1 or type 2 cytoplasmic tail is dominantly expressed; while in other chicken lines, both type 1 and type 2 are dominantly expressed.

7. BG genes might be associated with certain infectious diseases (Goto *et al.*, 2009). As described before, six BG genes out of eight found in T and B cells of line 6_1 chicken (B2) were identical to these BG genes typed in line CB chicken (B12) previously by Salomenson et al. (Salomonsen *et al.*, 2014). The equivalent BG genes in line CB were also detected in T cells and B cells, which fits the result of hemagglutination using anti-sera that chicken B2 and B12 haplotype have same BG antigens (Simonsen *et al.*, 1982). Therefore, in theory there should be some similar expression pattern observed between the two chicken lines. However, the two BG genes found in T cells of line CB were different from the three BG genes found in T cells of line 6_1. Although most (four out of five) BG genes found in B cells of line CB were also detected in B cells of line 6_1, one of the two dominantly expressed BG genes found in line CB was not observed in line 6_1. Most strikingly, the dominantly expressed BG gene in T cells of line 61, 6TBBGa, is identical to BG13 in line CB, but all the transcripts detected have intron retained after the first exon of cytoplasmic tail, resulting in very short tail. The cDNA sequences of 6TBBGa in B cells have a retained intron between exon 2 (encoding Ig-V domain) and exon 3 (encoding transmembrane region), potentially encode soluble BG proteins. One explanation would be that BG genes are associated with disease resistance: line CB and 6_1 have some identical BG genes originally, but during the selection for certain

diseases, the BG gene expression changed.

8. BG0 and BG1 might be important for BG functions in T cells and B cells. When the project was designed, BG0 and BG1 were not considered as objectives to be examined for two reasons. One reason was that BG0 was not detected in T cells or B cells during previous work on line CB which might be due to the primers that were universal for other BG genes but not for BG0 (Salomonsen *et al.*, 2014). The other reason was that BG0 and BG1 are the two singleton BG genes outside of BG region, and we thought they might not be involved with other BG genes in BG region. However, the cDNA sequences of both BG0 and BG1 were detected in this project, although further cloning and sequencing was not followed up, there is a possibility that BG0 and/or BG1 would form a heterodimer with other BG genes in T cells and B cells, which is worth testing in the future.

9. Additional information may leave some clues for future explorations. The three potential function domains, Ig-V domain, transmembrane region and cytoplasmic tail region, all have certain patterns which deserve further research. For example, the variation of amino acids within the BG genes in the same clade of phylogenetic tree built on nucleotide sequences of Ig-V domains is mostly located in the loop region, which might not influence the ligand binding. However, the variation between BG genes from different clades of the Ig-V domain phylogenetic tree is not only located in loop region and β-strands, but more strikingly, they all have their amino acid side chain pointing out of the Ig-V structure. This strongly suggests that BG genes from different clades of Ig-V domain phylogenetic tree bind to different ligands, or bind to the same ligands but in a different angel. Also, the Ig-V domain showed great polymorphism in the commercial chickens (Iglesias *et al.*, 2007), suggesting the complexities of function for Ig-V domains.

As for the transmembrane region, there are two types of transmembrane regions, and the major difference between the two types is that one type has a charged amino acid ‘K’. It has been known that quite a few immune receptors, both in innate and adaptive immune systems, bind to DAP signaling molecules to transduce intracellular signal. The basis of pairing DAP molecules is that there is a positive charged amino acid (usually R or K) at the same location in their transmembrane regions (Lanier, 2009). Take human killer-cell immunoglobulin-like receptors (KIRs) as examples, all the activating KIRs have a ‘K’ in the transmembrane region to bind DAP12 while inhibitory KIRs don’t (Campbell *et al.*, 2011). BG genes might have similar mechanism, and the alternatively spliced transcripts resulting in various lengths of

truncated cytoplasmic tails might be involved in the interaction between BG molecule and different adaptor molecules.

Last, there are two types of cytoplasmic tails, type 1 and type 2 with type 2 including two sub types, type 2a and type 2b. With the understanding of true organization of α-helical coiled coil, potential motifs would be more easily displayed and observed. However, further experiments need to be done to test these hypothesizes.

10. Useful tools for future studies on BG functions were established. Fourteen stable cell lines yielding large quantities of BG-Fc fusion proteins were developed, which are composed of the Ig-V domain of 14 BG genes from B12 haplotype (Salomonsen *et al.*, 2014), followed by a linker peptide and human IgG Fc fragment. BG-Fc fusion proteins themselves are powerful tool for functional study, and they were used to screen and characterize 290 BG mAb supernatants. In total 63 BG mAb supernatants were found specifically recognizing certain patterns of BG-Fc fusion proteins using sandwich ELISA. The BG mAbs would be used in future research for further characterizing BG proteins, e.g. tissue distribution, as well as BG functions.

7.2 Future work

With the revelation of many BG cDNA sequences and the development of essential tools for BG research, the plan for future work includes but is not limited to the following.

1. Designing an universal primer pair to amplify tissue BG genes. BG genes are known to be highly polymorphic, and due to the big difference in sequences between BG genes expressed in tissue and those by haemopoietic cells, it was unrealistic to design an universal primers to detect every BG gene. In this project, an universal primer pair called HU was developed and proved to be effective in amplifying BG cDNA sequences in T and B cells. An optimized protocol for fully understanding the alternative splicing transcripts was established, which set a good example for future work on tissue BG genes. Therefore, an universal primer pair amplifying tissue BG genes should be designed and applied on these important immune related tissue samples, for example gut, especially the intestinal epithelial cells, bursa, thymus etc.

2. Genomic typing of BG genes in order to understand the evolution history and associations

with diseases. With the development of sequencing techniques, genomic typing of BG genes in the BG region should be considered to fully understand the BG gene evolution; RNA-seq would help to identify the relationships between BG genes and diseases.

3. Further characterizing BG mAbs. Although the BG mAbs supernatants have been screened by 14 BG-Fc fusion protein of B12 haplotype, these specific mAbs should be further characterized. The hybridoma should be grown to single clone and tested by western blots using BG-Fc fusion proteins and by flow cytometry after staining with red blood cells of B12 haplotype chicken. Once the specificity was determined, the first experiment is to detect 'soluble BG proteins' in chicken B cell lines and see whether those alternative splicing transcripts having intron retained after exon 2 (Ig-V domain) could produce soluble BG proteins. Then the BG mAbs can be used to investigate the BG protein tissue distribution and other functional studies.

4. Proteomics study to search ligands for cytoplasmic tails of BG proteins. In this project, lots of alternative splicing isoforms were detected for most BG genes, and it would be most interesting to see to which molecules these isoforms bind and consider their functions.

5. Ligands searching for Ig-V domains of BG proteins and functional assay using BG-Fc fusion proteins. The BG-Fc fusion proteins could be used to stain lymphocytes extracted from blood and see whether BG proteins bind any cells and, if so, to try to find their ligands. Also the BG-Fc fusion proteins could be applied for functional assays, especially on γδT cells, to see if any BG genes could stimulate or inhibit T cell proliferation.

6. CRISPR knockout certain BG genes to understand BG functions. The CRISPR technique has been successfully used in our lab on other research. It would be usefully to knock out certain BG genes to investigate the roles of BG genes during infections in vivo.

Appendix A. Reagents and formulae

A1. Tissue culture reagents

1. Human embryonic kidney cells 293 (HEK293) and HEK293 derived stable cell lines were maintained in dulbeco's modified eagle's medium (DMEM) (Sigma) with 10% fetal bovine serum (FBS) (Gibco®) and antibiotics (penicillin 100 U/ml and streptomycin 100 µg/ml).

2. Transfection reagent was jetPEI in vitro DNA kit (Polyplus Transfection); selection antibody for making stable cell lines was Geneticin 418 (G418 for short) (Santa Cruz Biotechnology).

3. Cells were washed by Phosphate buffered saline (PBS), pH 7.4 (Sigma).

4. Cells were split using 0.05% Trypsin-EDTA solution (Gibco®).

A2. Molecular biology reagents and commercial kits

Lysogeny broth (LB) medium: 1% w/v Tryptone, 0.5% w/v Yeast Extract, 1% w/v NaCl
Note: For LB plate, 1% w/v Agar was added into LB medium; for LB-ampicillin or LB-ampicillin plates, 100 µg/mL ampicillin was added into LB medium or LB plate

TAE buffer: 40 mM Tris base, 20 mM acetic acid, and 1 mM ethylenediaminetetraacetic acid (EDTA) (pH 8.0)

TE buffer, pH 7.4: 10 mM Tris.Cl, and 1 mM EDTA

2x IEL buffer (100mL): 68 mL PBS (without calcium/magnesium), 20 ml fetal bovine serum (FBS), 2 ml 100 mM sodium pyruvate, 4 ml 1 M 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 4 ml 500 mM EDTA, 2 ml 10000 U/mL penicillin-streptomycin

Cloning vector: pJET1.2/blunt cloning vector (Thermo Scientific)

Competent cell: homemade *E. coli* DH5α

RNA extraction: NucleoSpin® RNA II RNA extraction kit (Machery-Nagel)

cDNA amplification: Maxima H Minus First Strand cDNA Synthesis Kit (ThermoFisher)

PCR reaction: Phusion® Hot Start Flex DNA Polymerase (NEB), Taq DNA Polymerase Recominant (Invitrogen), MyTaq™ Red Mix (Bioline)

Gel/PCR product purification: ISOLATE II PCR and Gel Kit (Bioline)

Miniprep & Midiprep: PureLink® Quick Plasmid Miniprep Kit (Invitrogen), QIAGEN® Plasmid Mini, Midi Kits (QIAGEN)

DNA ligation: T4 DNA ligase (NEB)

Blunt ligation: CloneJET PCR cloning kit (ThermoFisher)

TA cloning: TOPO® TA Cloning® kit (Invitrogen)

DNA site mutation: Q5 Site-Directed Mutagenesis kit (NEB)

A3. Biochemistry reagents and commercial kits

PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM K_2HPO_4

TBST: 140 mM NaCl, 50 mM TrisCl, 0.05% Tween-20 (pH 8.0)

SDS-PAGE separating gel 10% (per gel): 8.425 ml De-ionized water, 5.250 ml 1.5M Tris/HCl pH8.8, 0.210 ml 10% sodium dodecyl sulphate, 7.000 ml Acrylamide stock 30%, 0.105 ml Ammonium persulphate 10%, 0.011 ml TEMED

SDS-PAGE stacking gel 4% (per gel): 3.123 ml De-ionized water, 1.375 ml 1.5M Tris/HCl pH8.8, 0.055 ml 10% sodium dodecyl sulphate, 0.733 ml Acrylamide stock 30%, 0.028 ml Ammonium persulphate 10%, 0.003 ml TEMED

2x SDS-PAGE loading buffer : 2% sodium dodecyl sulphate (SDS), 20% glycerol, 20 mM Tris-Cl, 2 mM EDTA and 0.1 mg/ml Bromphenol Blue Dye
Note: for reduced gel, DTT is added to make a final concentration of 200 mM

10x SDS-PAGE running buffer (1L): 30 g Tris base, 144 g glycine, 10 g SDS

Western blot (WB) transfer buffer (1L): 3.01 g Tris base, 14.4 g glycine, 100 mL methanol

WB blocking buffer: 5-10% skimmed milk in TBST buffer.

Protein washing buffer: 0.1 M sodium phosphate, 0.15 M NaCl, pH7.4

Protein elution buffer: 0.2 M glycine, pH2.7

Protein G beads: Protein G Immobilized on Agarose (Sigma)

Appendix B. Primers used in this thesis

Primer name	Description	Sequence (5' to 3')
UC74	SS: signal sequence forward primer	CTCCTGCCTTATCTCRTGGCTCTGCAC
UC76	TM: transmembrane reverse primer	CACAGCCAGAGCCACYKTCCAG
UC206	H: Haemopoietic BG forward Primer	TCCGCTCGAGCTCTCTYCTCCTACAG
UC210	2: type 2 BG reverse primer	TTTCTACCTTTGYTTCSGCGTGCATG
UC453	BG internal forward sequencing primer	TGTKGTGYTGYGCTGCCA
UC454	BG internal forward sequencing primer	GAGCCTCCGTGTCACTGCCA
UC644	BG13-Fc reverse primer	CCGCTCGAGATCTGACACCTCCAGG
UC645	BG10-Mutation-Forward primer	CTGGAGGTGTCAGATCTCGAGAGCGGAGG
UC646	BG10-Mutation-Reverese primer	CCTCCGCTCTCGAGATCTGACACCTCCAG
UC647	BG internal reverse sequencing primer	GCAGTGTTCTCACTCAA
UC648	BG10-Fc Reverse Primer	CCGCTCGAGATCTGACACCTCCAGCTC
UC649	BG internal reverse sequencing primer	GCAGTTMATHTCTCARTC
UC650	BG-H Reverse Primer	TAACACCCAAAGCAGTTTTCTNCC
UC699	BG internal forward sequencing primer	TTTTCTATGATCATCC
UC700	BG internal forward sequencing primer	TTTTCTATGATCATCC
UC701	BG internal forward sequencing primer	TGGCTCTGCACYTCCTCS
UC702	BG internal forward sequencing primer	TGGCTCTGCACYTCCTCC
UC703	BG internal forward sequencing primer	TGRACCTGGAGGTGTCAG
T7	sequencing primer	TAATACGACTCACTATAGGG
pJETR	sequencing primer	AAGAACATCGATTTTCCATGGCAG
P1	SigpIg plus sequencing primer	CCGCTCTCGAGATCTGACAC
P2	SigpIg plus sequencing primer	CAATAGGGGGCGTACTTGGC
Fc	SigpIg plus sequencing primer	TGAGCCACGAAGACCCTGAGG

Appendix C.

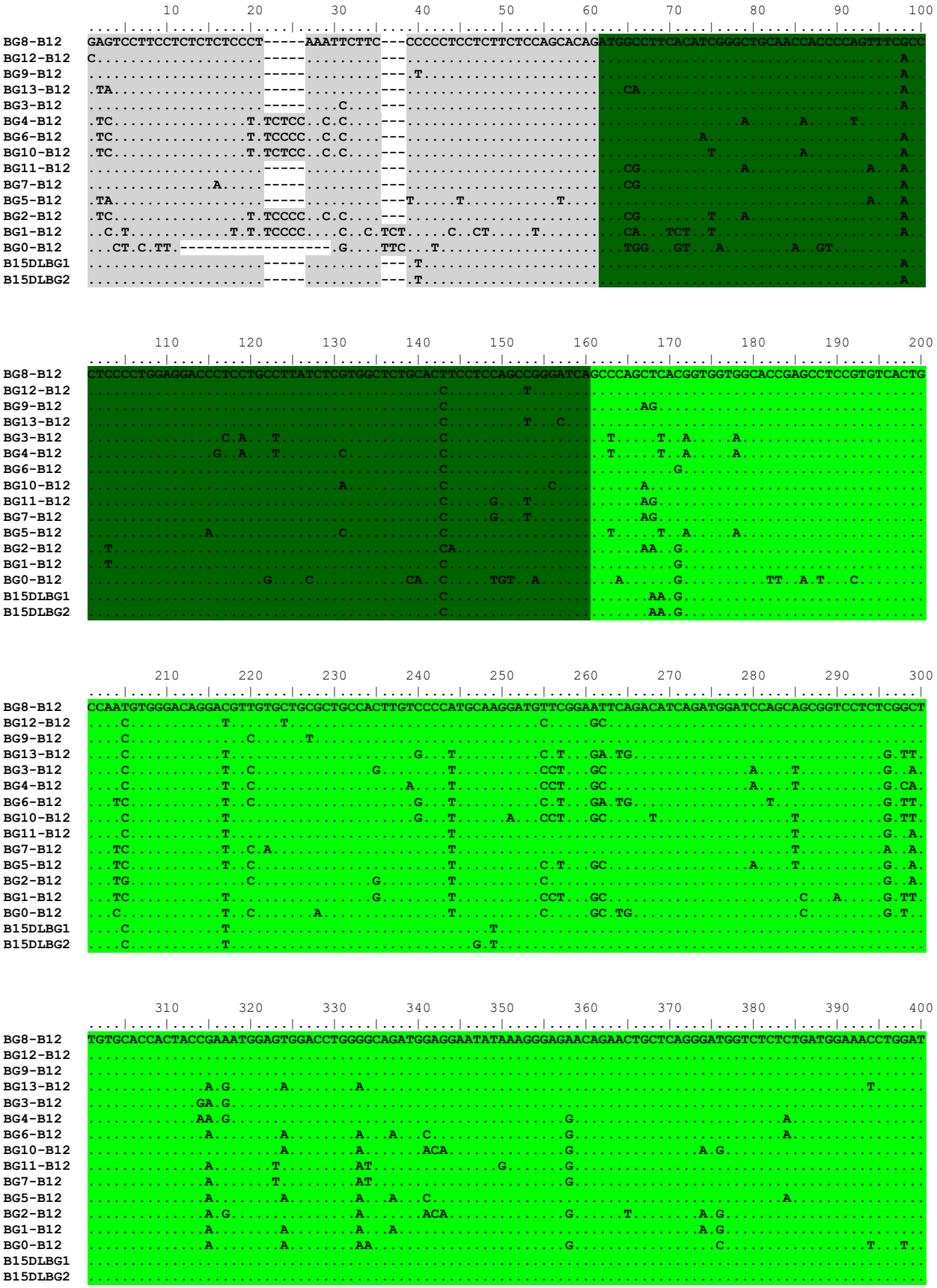


Figure legend is shown on the page following the figure

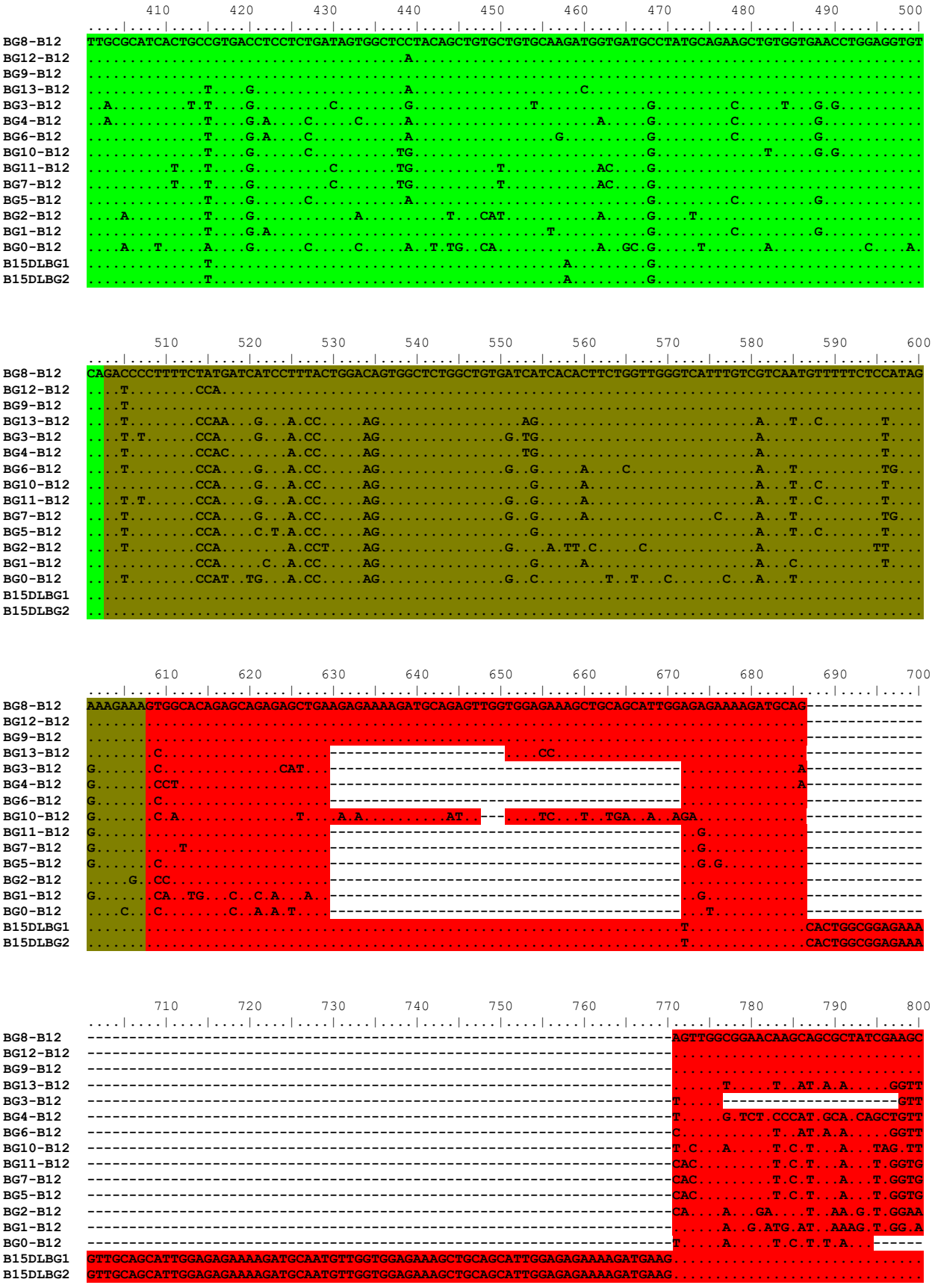


Figure legend is shown on the page following the figure

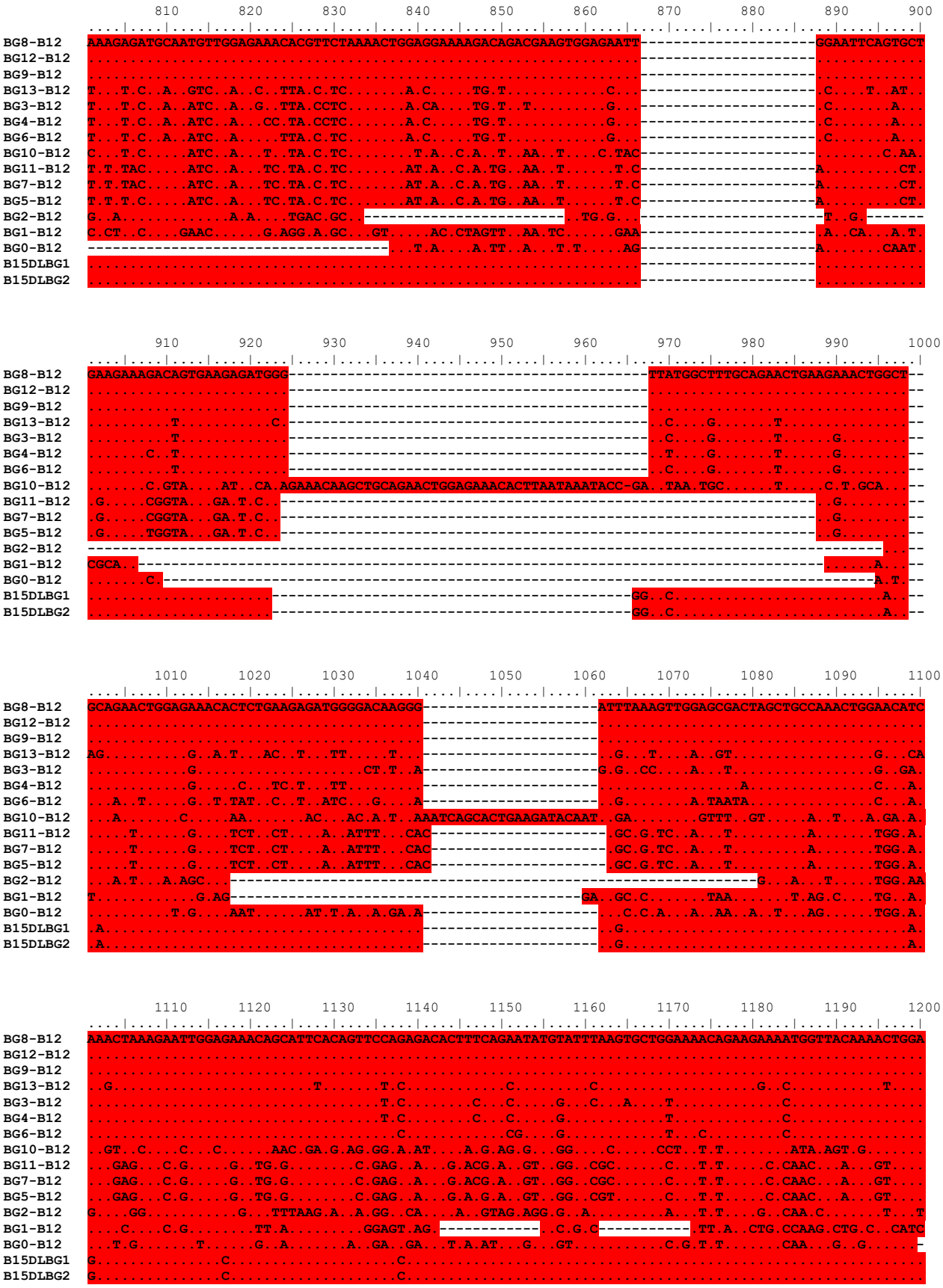


Figure legend is shown on the page following the figure

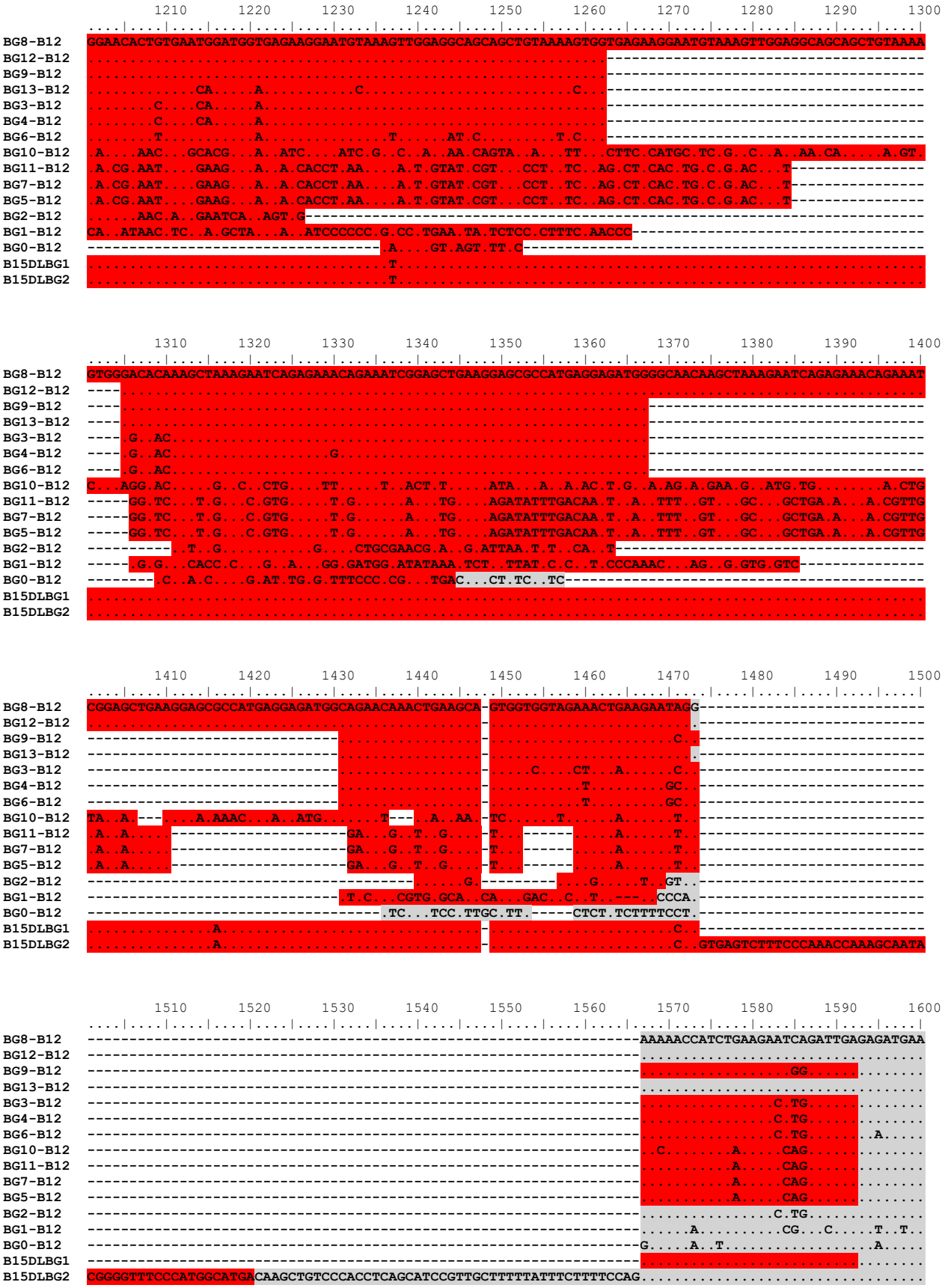
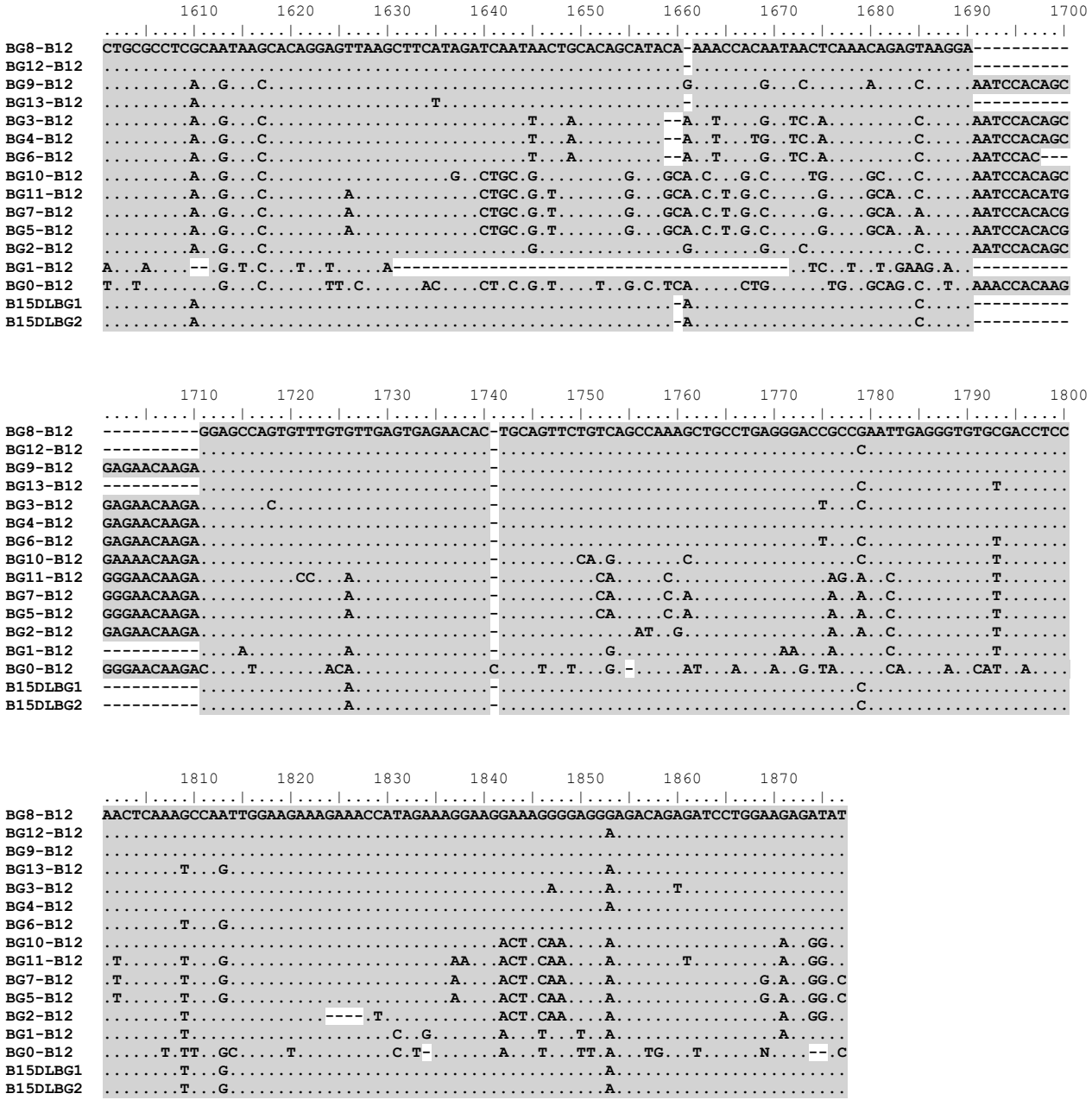
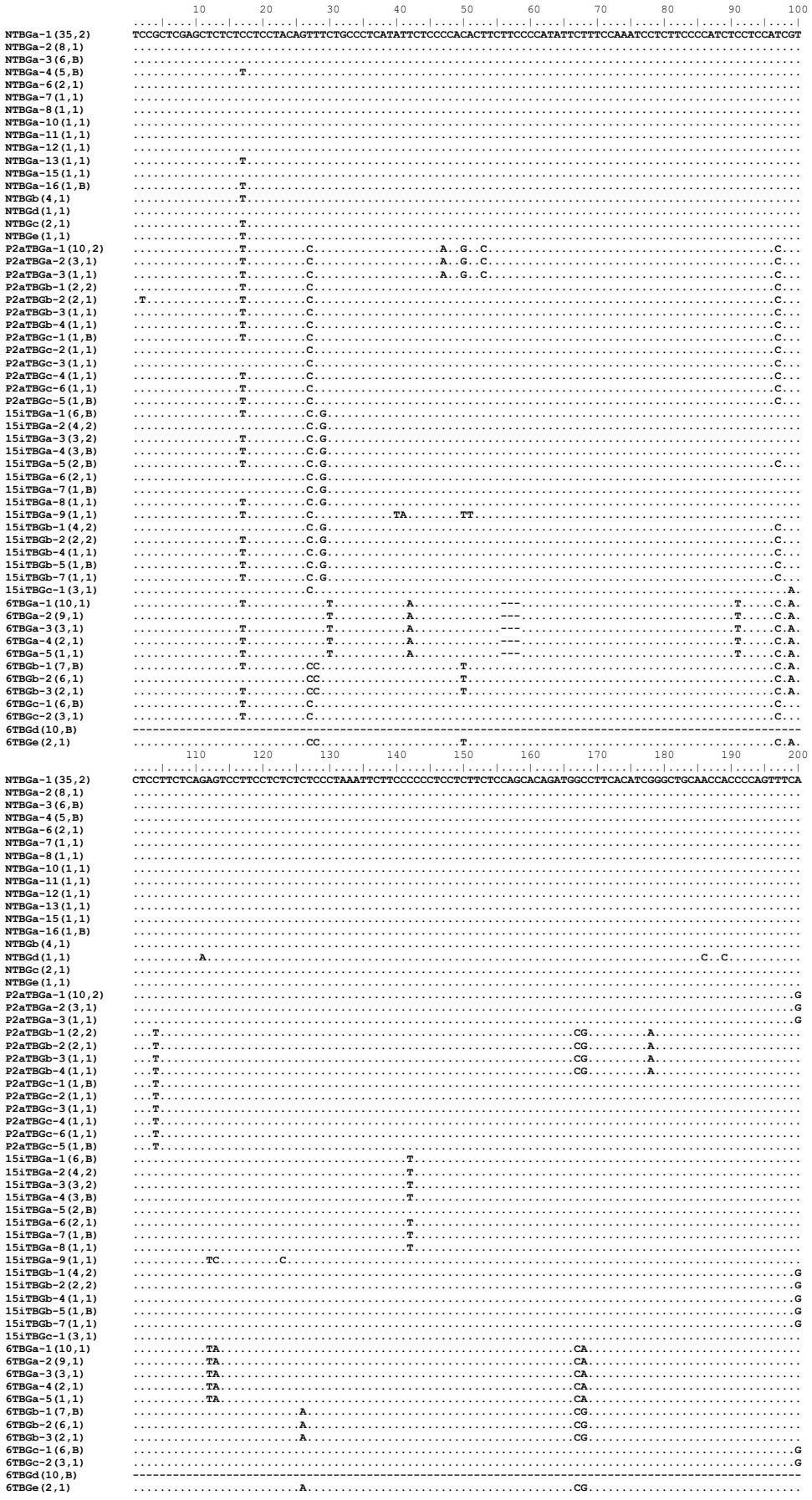


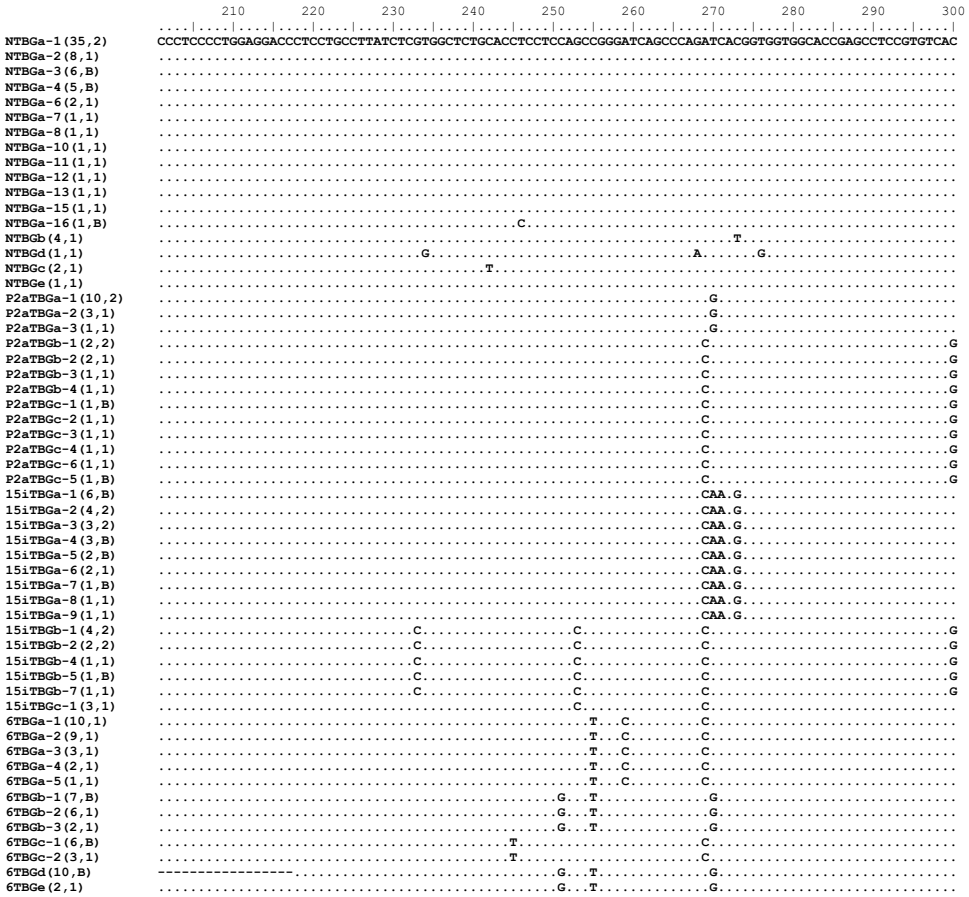
Figure legend is shown on the page following the figure

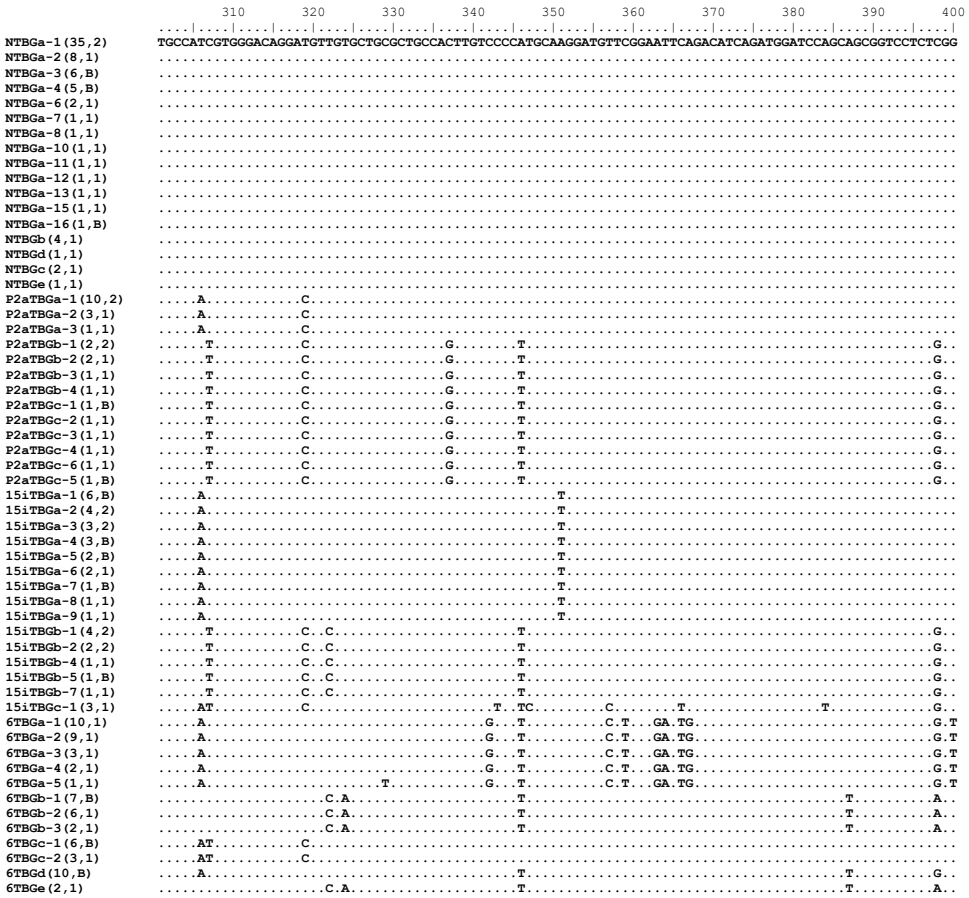


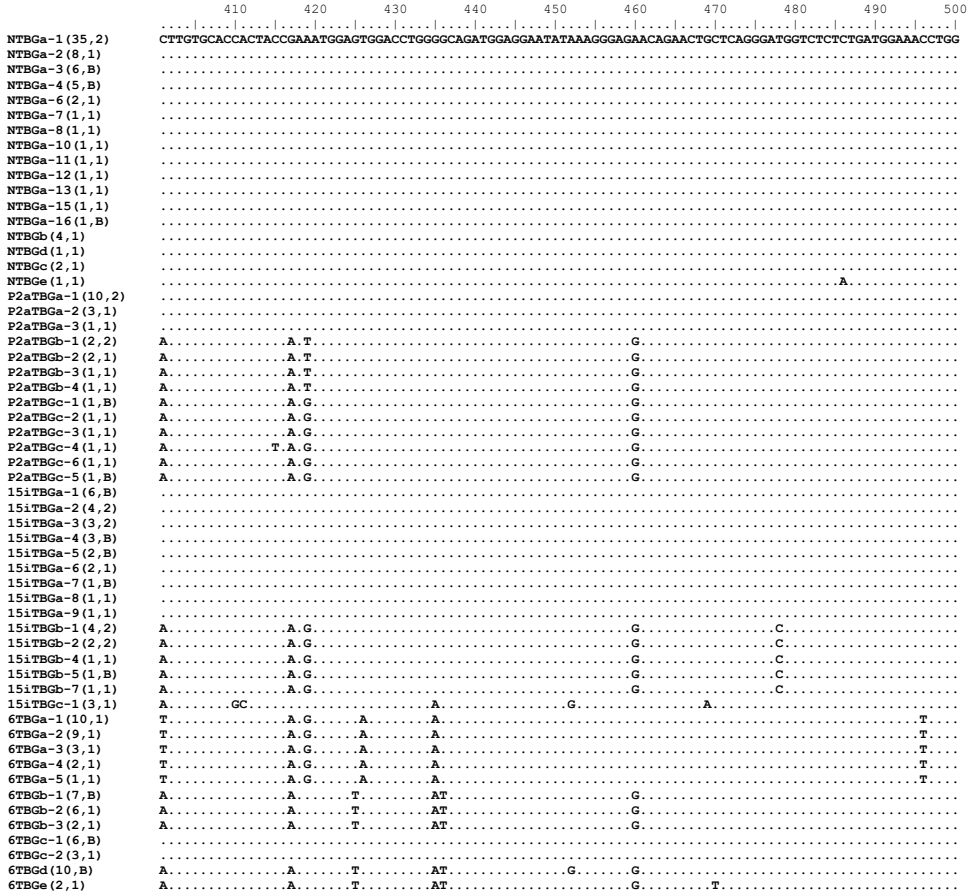
Appendix C. The cDNA sequence alignment between the two new BG genes found in duodenum sample of B15 haplotype using H2 primers and 14 BG genes from B12 haplotype. The top 14 sequences are BG genes from B12 haplotype and the bottom two sequences are the two new BG genes from duodenum sample of B15 haplotype. Colours indicate different coding regions: grey for 5' UTRs and 3' UTRs; dark green for signal sequences; light green for Ig-V domains; brown for transmembrane regions; red for cytoplasmic tail regions. Letters indicate nucleotides, dots indicate identity with the top sequence (BG8 from B12 haplotype), and dashes indicate no sequence present compared to one or more of the other sequences.

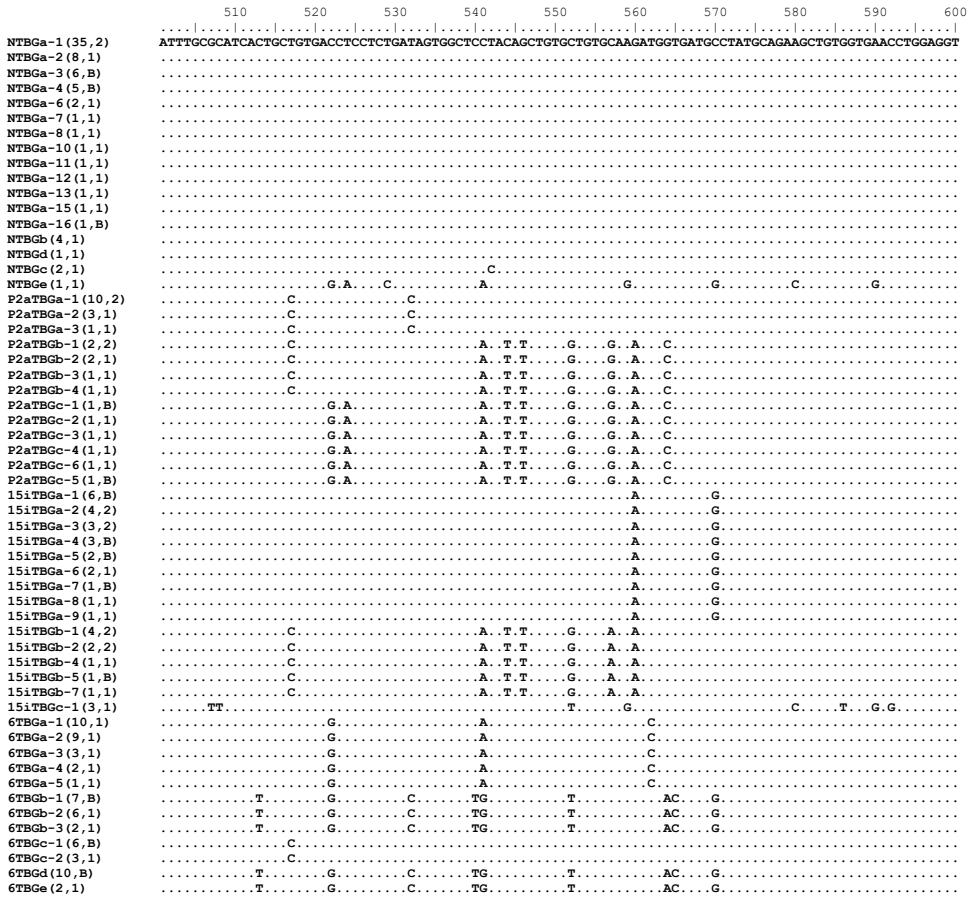
Appendix D.

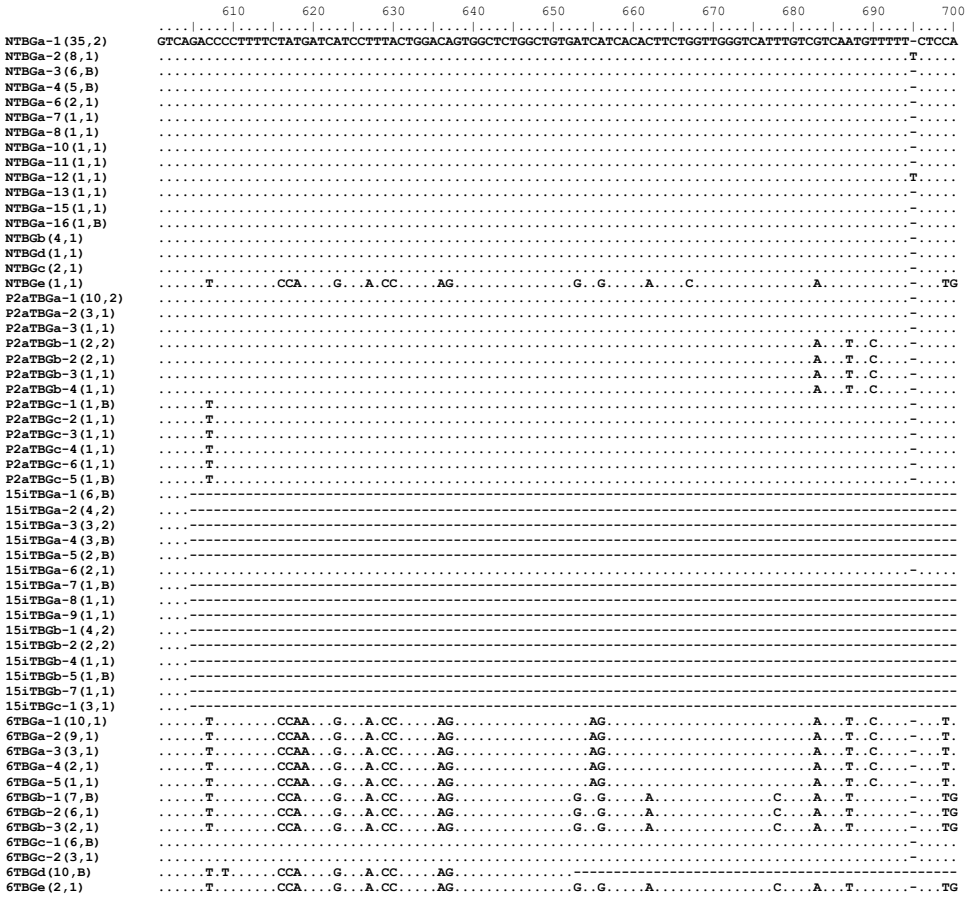

10 20 30 40 50 60 70 80 90 100
NTBGa-1(35,2) TCCGCTCGAGCTCTCTCCTCCTACAGTTTCTGCCCTCATATTCTCCCCACACTTCTTCCCCATATTCTTTCCAAATCCTCTTCCCCATCTCCTCCATCGT
110 120 130 140 150 160 170 180 190 200
NTBGa-1(35,2) CTCCTTCTCAGAGTCCTTCCTCTCTCTCCCTAAATTCTTCCCCCCTCCTCTTCTCCAGCACAGATGGCCTTCACATCGGGCTGCAACCACCCCAGTTTCA

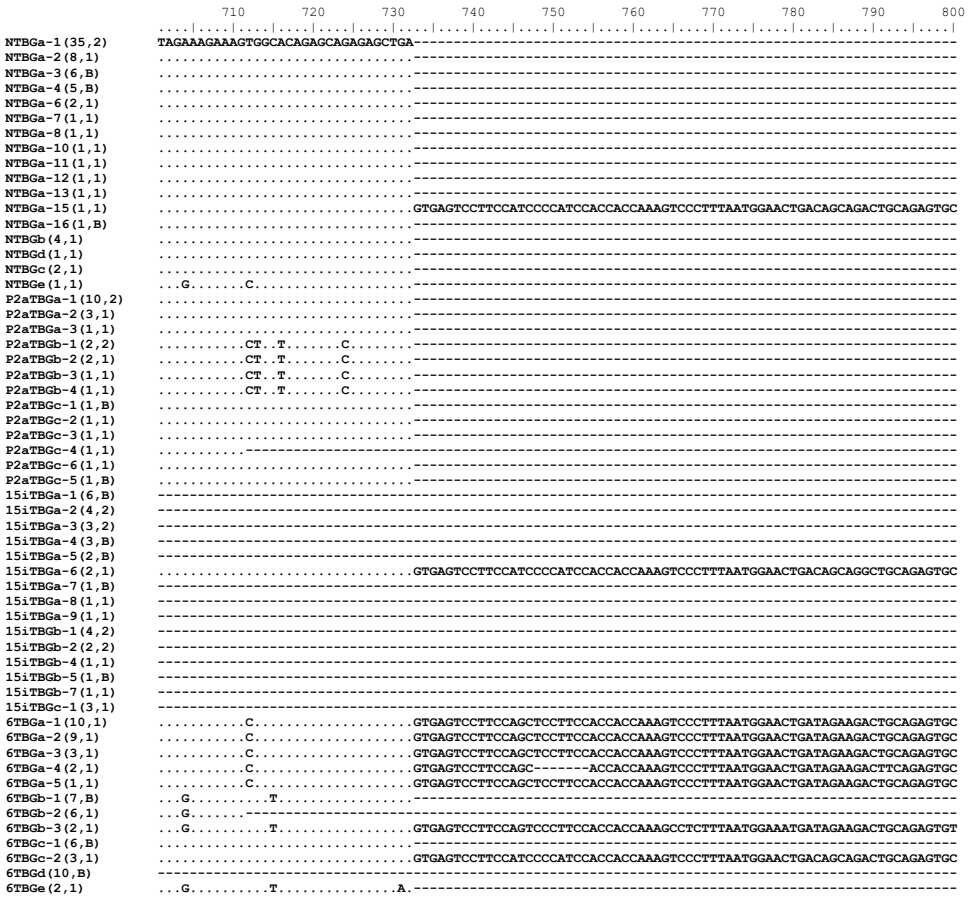

210 220 230 240 250 260 270 280 290 300
NTBGa-1(35,2) CCCTCCCCTGGAGGACCCTCCTGCCTTATCTCGTGGCTCTGCACCTCCTCCAGCCGGGATCAGCCCAGATCACGGTGGTGGCACCGAGCCTCCGTGTCAC
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

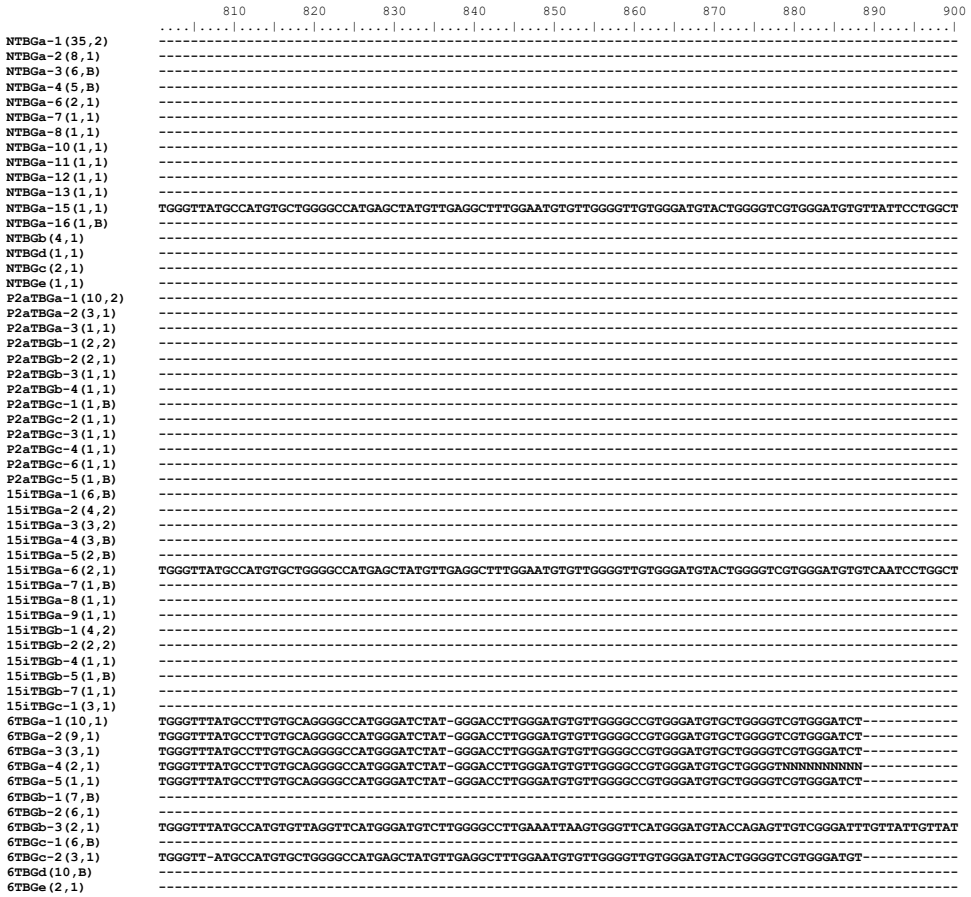

310 320 330 340 350 360 370 380 390 400
NTBGa-1(35,2) TGCCATCGTGGGACAGGATGTTGTGCTGCGCTGCCACTTGTCCCCATGCAAGGATGTTCGGAATTCAGACATCAGATGGATCCAGCAGCGGTCCTCTCGG
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

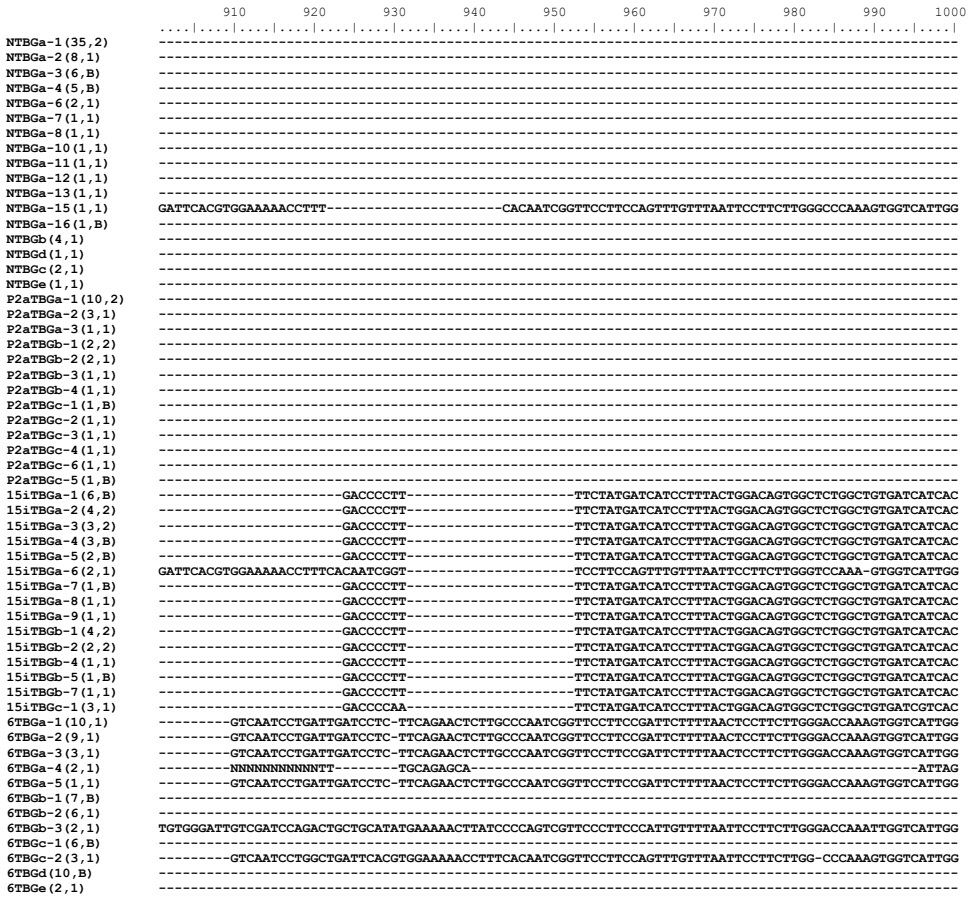

410 420 430 440 450 460 470 480 490 500
NTBGa-1(35,2) CTTGTGCACCACTACCGAAATGGAGTGGACCTGGGGCAGATGGAGGAATATAAAGGGAGAACAGAACTGCTCAGGGATGGTCTCTCTGATGGAAACCTGG
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

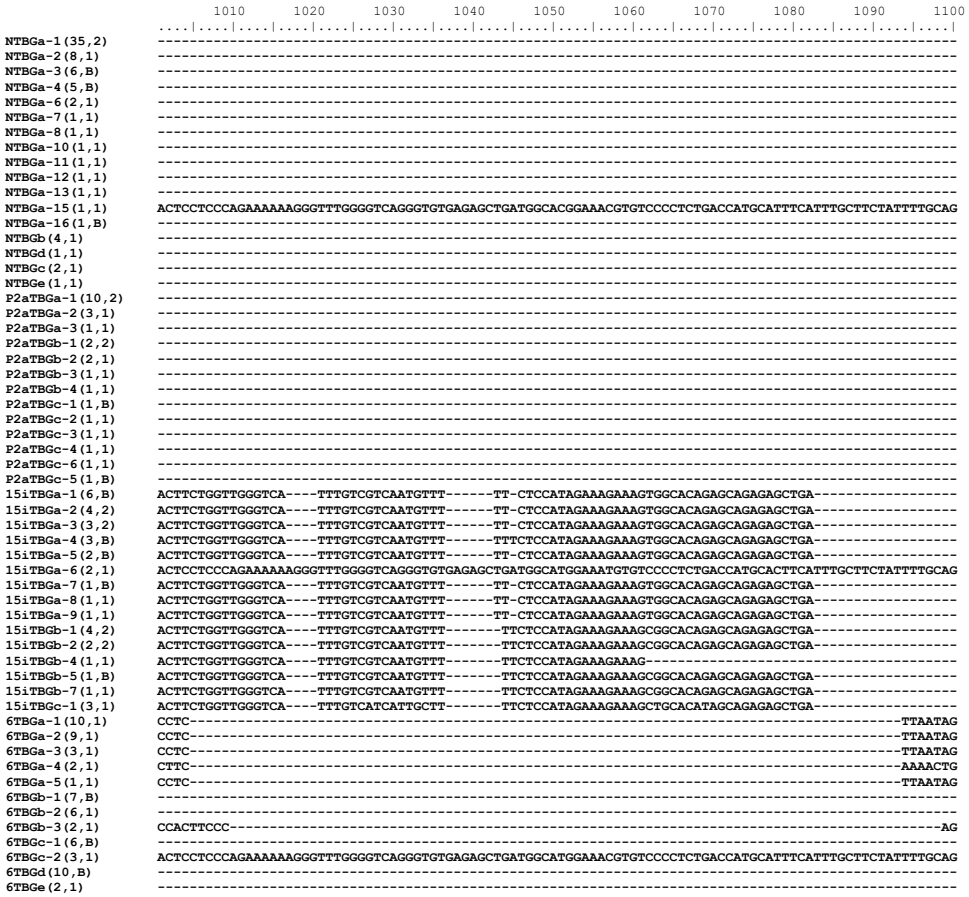

510 520 530 540 550 560 570 580 590 600
NTBGa-1(35,2) ATTTGCGCATCACTGCTGTGACCTCCTCTGATAGTGGCTCCTACAGCTGTGCTGTGCAAGATGGTGATGCCTATGCAGAAGCTGTGGTGAACCTGGAGGT
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

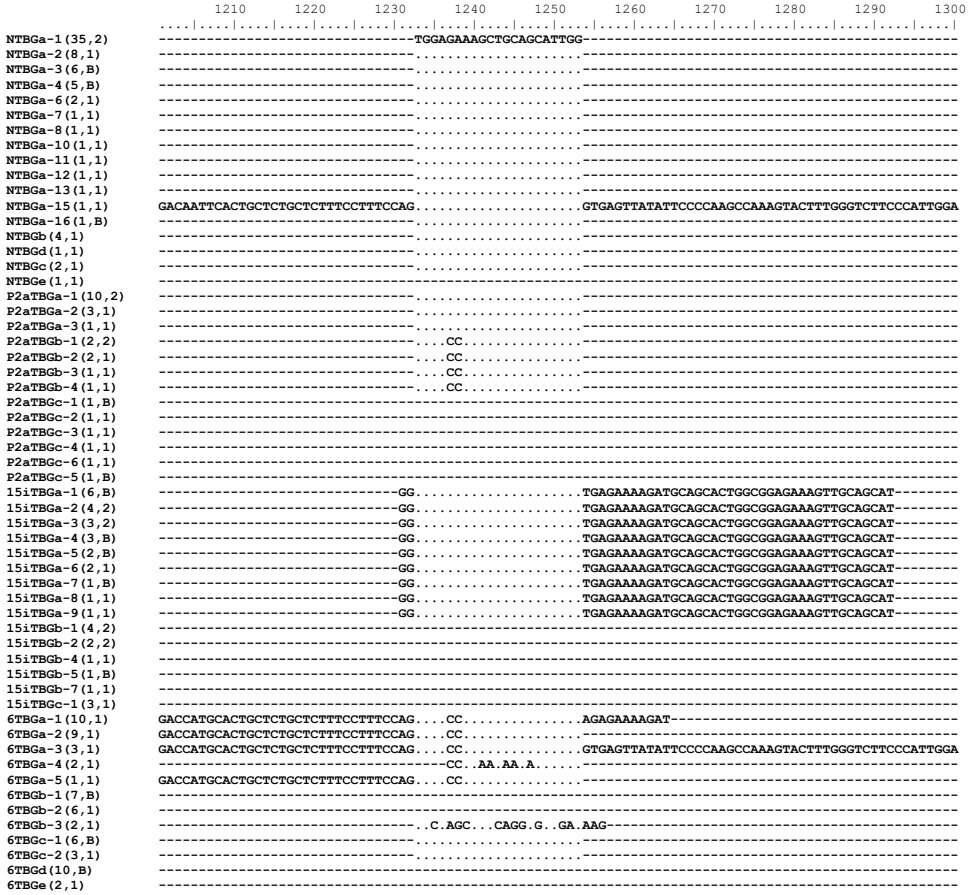

610 620 630 640 650 660 670 680 690 700
NTBGa-1(35,2) GTCAGACCCCTTTTCTATGATCATCCTTTACTGGACAGTGGCTCTGGCTGTGATCATCACACTTCTGGTTGGGTCATTTGTCGTCAATGTTTTT-CTCCA
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

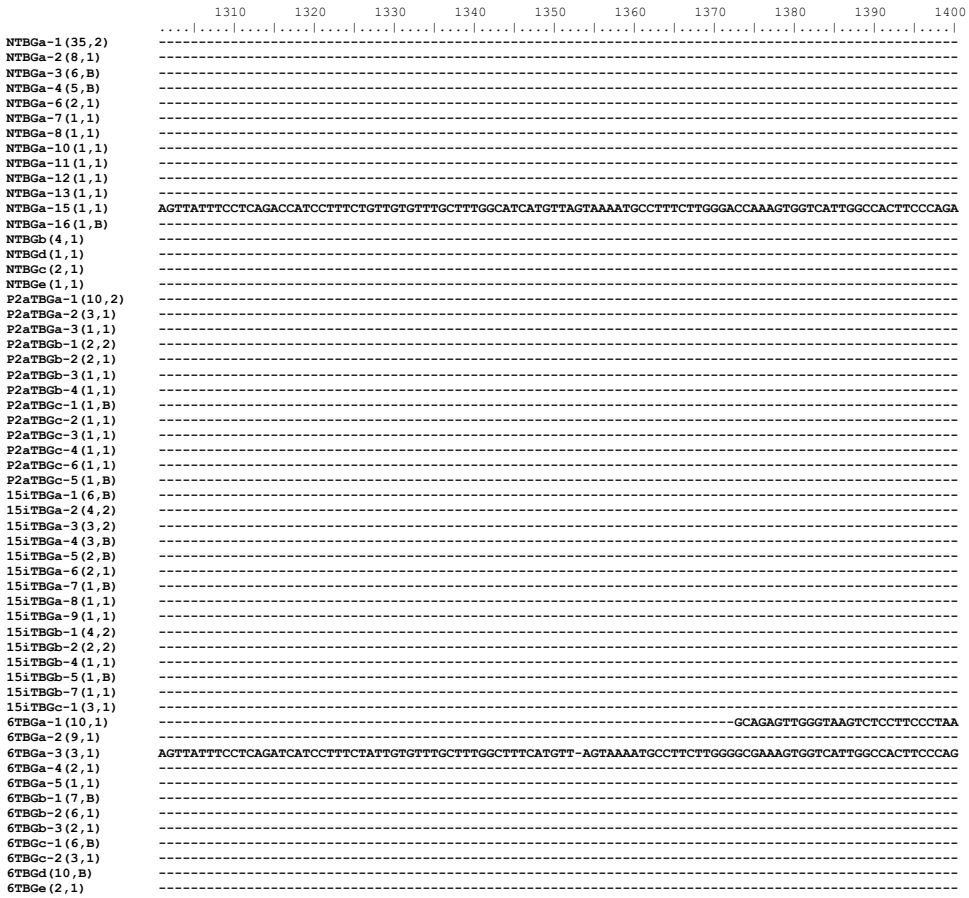

710 720 730 740 750 760 770 780 790 800
NTBGa-1(35,2) TAGAAAGAAAGTGGCACAGAGCAGAGAGCTGA
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1) GTGAGTCCTTCCATCCCCATCCACCACCAAAGTCCCTTTAATGGAACTGACAGCAGACTGCAGAGTGC
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1) GTGAGTCCTTCCATCCCCATCCACCACCAAAGTCCCTTTAATGGAACTGACAGCAGGCTGCAGAGTGC
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1) GTGAGTCCTTCCAGCTCCTTCCACCACCAAAGTCCCTTTAATGGAACTGATAGAAGACTGCAGAGTGC
6TBGa-2(9,1) GTGAGTCCTTCCAGCTCCTTCCACCACCAAAGTCCCTTTAATGGAACTGATAGAAGACTGCAGAGTGC
6TBGa-3(3,1) GTGAGTCCTTCCAGCTCCTTCCACCACCAAAGTCCCTTTAATGGAACTGATAGAAGACTGCAGAGTGC
6TBGa-4(2,1) GTGAGTCCTTCCAGC-------ACCACCAAAGTCCCTTTAATGGAACTGATAGAAGACTTCAGAGTGC
6TBGa-5(1,1) GTGAGTCCTTCCAGCTCCTTCCACCACCAAAGTCCCTTTAATGGAACTGATAGAAGACTGCAGAGTGC
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1) GTGAGTCCTTCCAGTCCCTTCCACCACCAAAGCCTCTTTAATGGAAATGATAGAAGACTGCAGAGTGT
6TBGc-1(6,B)
6TBGc-2(3,1) GTGAGTCCTTCCATCCCCATCCACCACCAAAGTCCCTTTAATGGAACTGACAGCAGACTGCAGAGTGC
6TBGd(10,B)
6TBGe(2,1)

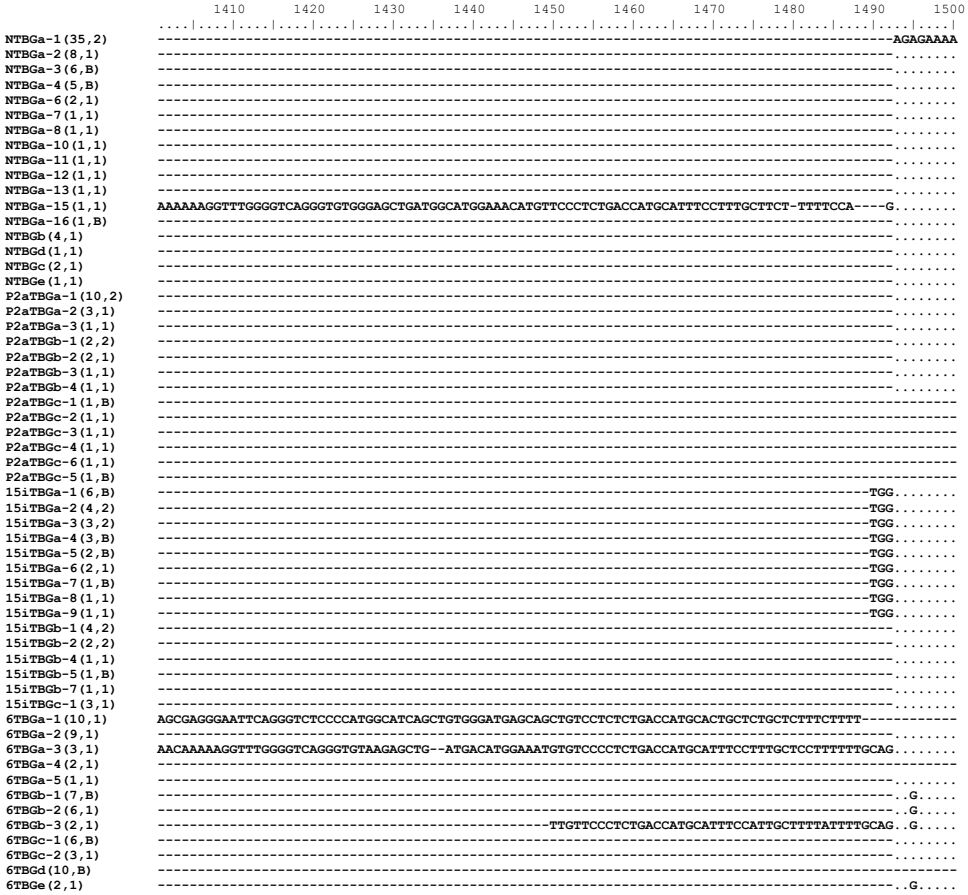

810 820 830 840 850 860 870 880 890 900
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1) TGGGTTATGCCATGTGCTGGGGCCATGAGCTATGTTGAGGCTTTGGAATGTGTTGGGGTTGTGGGATGTACTGGGGTCGTGGGATGTGTTATTCCTGGCT
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1) TGGGTTATGCCATGTGCTGGGGCCATGAGCTATGTTGAGGCTTTGGAATGTGTTGGGGTTGTGGGATGTACTGGGGTCGTGGGATGTGTCAATCCTGGCT
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1) TGGGTTTATGCCTTGTGCAGGGGCCATGGGATCTAT-GGGACCTTGGGATGTGTTGGGGCCGTGGGATGTGCTGGGGTCGTGGGATCT----------
6TBGa-2(9,1) TGGGTTTATGCCTTGTGCAGGGGCCATGGGATCTAT-GGGACCTTGGGATGTGTTGGGGCCGTGGGATGTGCTGGGGTCGTGGGATCT----------
6TBGa-3(3,1) TGGGTTTATGCCTTGTGCAGGGGCCATGGGATCTAT-GGGACCTTGGGATGTGTTGGGGCCGTGGGATGTGCTGGGGTCGTGGGATCT----------
6TBGa-4(2,1) TGGGTTTATGCCTTGTGCAGGGGCCATGGGATCTAT-GGGACCTTGGGATGTGTTGGGGCCGTGGGATGTGCTGGGGTNNNNNNNNNN----------
6TBGa-5(1,1) TGGGTTTATGCCTTGTGCAGGGGCCATGGGATCTAT-GGGACCTTGGGATGTGTTGGGGCCGTGGGATGTGCTGGGGTCGTGGGATCT----------
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1) TGGGTTTATGCCATGTGTTAGGTTCATGGGATGTCTTGGGGCCTTGAAATTAAGTGGGTTCATGGGATGTACCAGAGTTGTCGGGATTTGTTATTGTTAT
6TBGc-1(6,B)
6TBGc-2(3,1) TGGGTT-ATGCCATGTGCTGGGGCCATGAGCTATGTTGAGGCTTTGGAATGTGTTGGGGTTGTGGGATGTACTGGGGTCGTGGGATGT----------
6TBGd(10,B)
6TBGe(2,1)

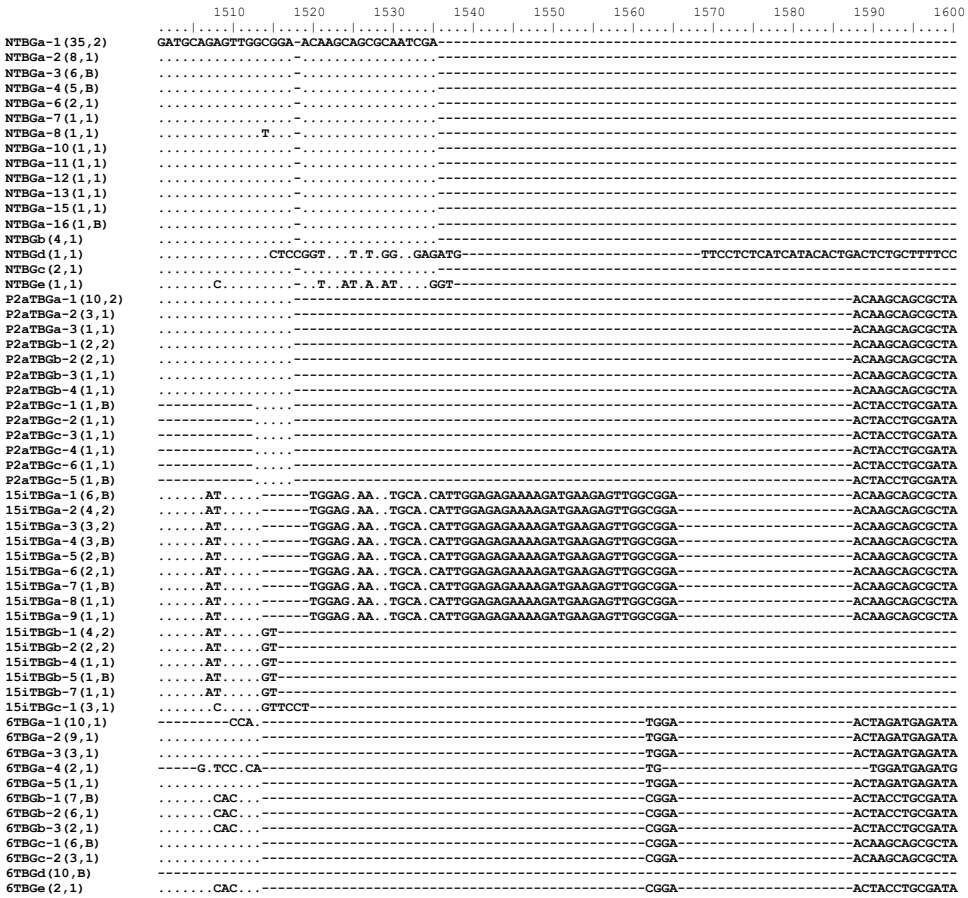

910 920 930 940 950 960 970 980 990 1000
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1) GATTCACGTGGAAAAACCTTT----------CACAATCGGTTCCTTCCAGTTTGTTTAATTCCTTCTTGGGCCCAAAGTGGTCATTGG
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B) ----------GACCCCTT----------TTCTATGATCATCCTTTACTGGACAGTGGCTCTGGCTGTGATCATCAC
15iTBGa-2(4,2) ----------GACCCCTT----------TTCTATGATCATCCTTTACTGGACAGTGGCTCTGGCTGTGATCATCAC
15iTBGa-3(3,2) ----------GACCCCTT----------TTCTATGATCATCCTTTACTGGACAGTGGCTCTGGCTGTGATCATCAC
15iTBGa-4(3,B) ----------GACCCCTT----------TTCTATGATCATCCTTTACTGGACAGTGGCTCTGGCTGTGATCATCAC
15iTBGa-5(2,B) ----------GACCCCTT----------TTCTATGATCATCCTTTACTGGACAGTGGCTCTGGCTGTGATCATCAC
15iTBGa-6(2,1) GATTCACGTGGAAAAACCTTTCACAATCGGT----------TCCTTCCAGTTTGTTTAATTCCTTCTTGGGTCCAAA-GTGGTCATTGG
15iTBGa-7(1,B) ----------GACCCCTT----------TTCTATGATCATCCTTTACTGGACAGTGGCTCTGGCTGTGATCATCAC
15iTBGa-8(1,1) ----------GACCCCTT----------TTCTATGATCATCCTTTACTGGACAGTGGCTCTGGCTGTGATCATCAC
15iTBGa-9(1,1) ----------GACCCCTT----------TTCTATGATCATCCTTTACTGGACAGTGGCTCTGGCTGTGATCATCAC
15iTBGb-1(4,2) ----------GACCCCTT----------TTCTATGATCATCCTTTACTGGACAGTGGCTCTGGCTGTGATCATCAC
15iTBGb-2(2,2) ----------GACCCCTT----------TTCTATGATCATCCTTTACTGGACAGTGGCTCTGGCTGTGATCATCAC
15iTBGb-4(1,1) ----------GACCCCTT----------TTCTATGATCATCCTTTACTGGACAGTGGCTCTGGCTGTGATCATCAC
15iTBGb-5(1,B) ----------GACCCCTT----------TTCTATGATCATCCTTTACTGGACAGTGGCTCTGGCTGTGATCATCAC
15iTBGb-7(1,1) ----------GACCCCTT----------TTCTATGATCATCCTTTACTGGACAGTGGCTCTGGCTGTGATCATCAC
15iTBGc-1(3,1) ----------GACCCCAA----------TTCTATGATCATCCTTTACTGGACAGTGGCTCTGGCTGTGATCGTCAC
6TBGa-1(10,1) ---------GTCAATCCTGATTGATCCTC-TTCAGAACTCTTGCCCAATCGGTTCCTTCCGATTCTTTTAACTCCTTCTTGGGACCAAAGTGGTCATTGG
6TBGa-2(9,1) ---------GTCAATCCTGATTGATCCTC-TTCAGAACTCTTGCCCAATCGGTTCCTTCCGATTCTTTTAACTCCTTCTTGGGACCAAAGTGGTCATTGG
6TBGa-3(3,1) ---------GTCAATCCTGATTGATCCTC-TTCAGAACTCTTGCCCAATCGGTTCCTTCCGATTCTTTTAACTCCTTCTTGGGACCAAAGTGGTCATTGG
6TBGa-4(2,1) ---------NNNNNNNNNNNTT--------TGCAGAGCA----------ATTAG
6TBGa-5(1,1) ---------GTCAATCCTGATTGATCCTC-TTCAGAACTCTTGCCCAATCGGTTCCTTCCGATTCTTTTAACTCCTTCTTGGGACCAAAGTGGTCATTGG
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1) TGTGGGATTGTCGATCCAGACTGCTGCATATGAAAAACTTATCCCCAGTCGTTCCCTTCCCATTGTTTTAATTCCTTCTTGGGACCAAATTGGTCATTGG
6TBGc-1(6,B)
6TBGc-2(3,1) ---------GTCAATCCTGGCTGATTCACGTGGAAAAACCTTTCACAATCGGTTCCTTCCAGTTTGTTTAATTCCTTCTTGG-CCCAAAGTGGTCATTGG
6TBGd(10,B)
6TBGe(2,1)

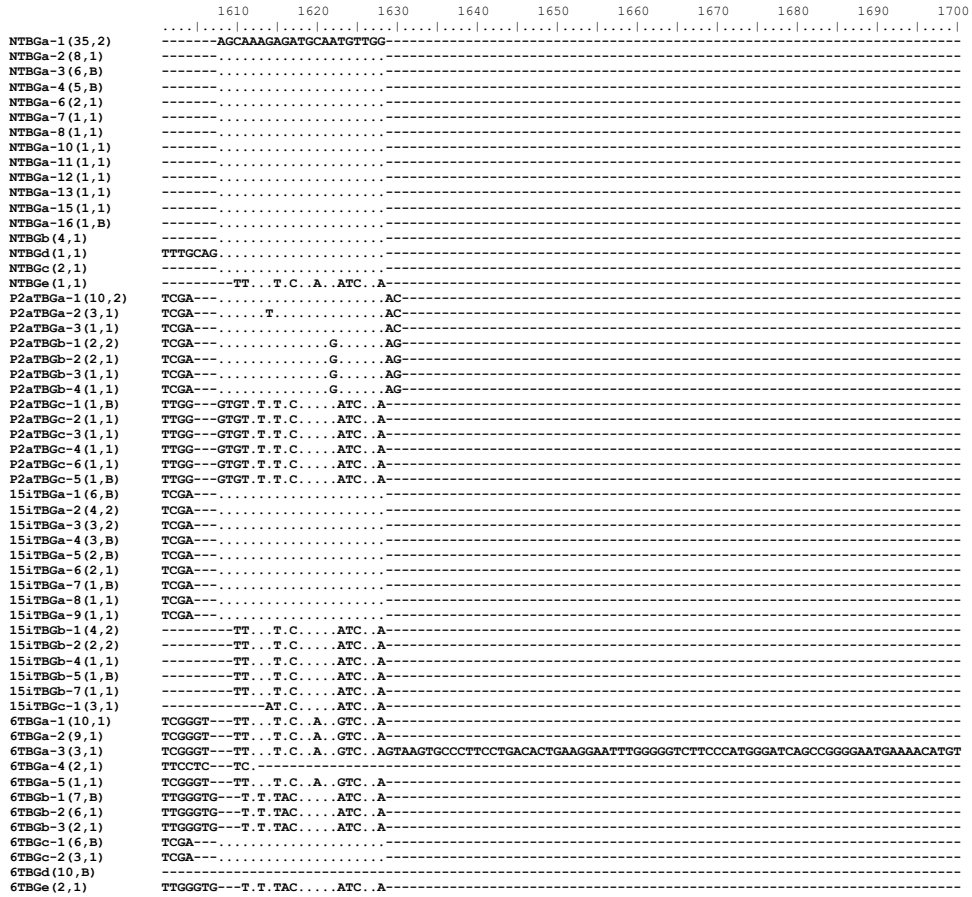

1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1) ACTCCTCCCAGAAAAAAGGGTTTGGGGTCAGGGTGTGAGAGCTGATGGCACGGAAACGTGTCCCCTCTGACCATGCATTTCATTTGCTTCTATTTTGCAG
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B) ACTTCTGGTTGGGTCA----TTTGTCGTCAATGTTT-----TT-CTCCATAGAAAGAAAGTGGCACAGAGCAGAGAGCTGA----------
15iTBGa-2(4,2) ACTTCTGGTTGGGTCA----TTTGTCGTCAATGTTT-----TT-CTCCATAGAAAGAAAGTGGCACAGAGCAGAGAGCTGA----------
15iTBGa-3(3,2) ACTTCTGGTTGGGTCA----TTTGTCGTCAATGTTT-----TT-CTCCATAGAAAGAAAGTGGCACAGAGCAGAGAGCTGA----------
15iTBGa-4(3,B) ACTTCTGGTTGGGTCA----TTTGTCGTCAATGTTT-----TTTCTCCATAGAAAGAAAGTGGCACAGAGCAGAGAGCTGA----------
15iTBGa-5(2,B) ACTTCTGGTTGGGTCA----TTTGTCGTCAATGTTT-----TT-CTCCATAGAAAGAAAGTGGCACAGAGCAGAGAGCTGA----------
15iTBGa-6(2,1) ACTCCTCCCAGAAAAAAGGGTTTGGGGTCAGGGTGTGAGAGCTGATGGCATGGAAATGTGTCCCCTCTGACCATGCACTTCATTTGCTTCTATTTTGCAG
15iTBGa-7(1,B) ACTTCTGGTTGGGTCA----TTTGTCGTCAATGTTT-----TT-CTCCATAGAAAGAAAGTGGCACAGAGCAGAGAGCTGA----------
15iTBGa-8(1,1) ACTTCTGGTTGGGTCA----TTTGTCGTCAATGTTT-----TT-CTCCATAGAAAGAAAGTGGCACAGAGCAGAGAGCTGA----------
15iTBGa-9(1,1) ACTTCTGGTTGGGTCA----TTTGTCGTCAATGTTT-----TT-CTCCATAGAAAGAAAGTGGCACAGAGCAGAGAGCTGA----------
15iTBGb-1(4,2) ACTTCTGGTTGGGTCA----TTTGTCGTCAATGTTT-------TTCTCCATAGAAAGAAAGCGGCACAGAGCAGAGAGCTGA----------
15iTBGb-2(2,2) ACTTCTGGTTGGGTCA----TTTGTCGTCAATGTTT-------TTCTCCATAGAAAGAAAGCGGCACAGAGCAGAGAGCTGA----------
15iTBGb-4(1,1) ACTTCTGGTTGGGTCA----TTTGTCGTCAATGTTT-------TTCTCCATAGAAAGAAAG----------
15iTBGb-5(1,B) ACTTCTGGTTGGGTCA----TTTGTCGTCAATGTTT-------TTCTCCATAGAAAGAAAGCGGCACAGAGCAGAGAGCTGA----------
15iTBGb-7(1,1) ACTTCTGGTTGGGTCA----TTTGTCGTCAATGTTT-------TTCTCCATAGAAAGAAAGCGGCACAGAGCAGAGAGCTGA----------
15iTBGc-1(3,1) ACTTCTGGTTGGGTCA----TTTGTCATCATTGCTT-------TTCTCCATAGAAAGAAAGCTGCACATAGCAGAGAGCTGA----------
6TBGa-1(10,1) CCTC----------TTAATAG
6TBGa-2(9,1) CCTC----------TTAATAG
6TBGa-3(3,1) CCTC----------TTAATAG
6TBGa-4(2,1) CTTC----------AAAACTG
6TBGa-5(1,1) CCTC----------TTAATAG
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1) CCACTTCCC----------AG
6TBGc-1(6,B)
6TBGc-2(3,1) ACTCCTCCCAGAAAAAAGGGTTTGGGGTCAGGGTGTGAGAGCTGATGGCATGGAAACGTGTCCCCTCTGACCATGCATTTCATTTGCTTCTATTTTGCAG
6TBGd(10,B)
6TBGe(2,1)
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
NTBGa-1(35,2) AGAGAAAAGATGCAGAGTTGG----------
NTBGa-2(8,1) .....----------
NTBGa-3(6,B) .....----------
NTBGa-4(5,B) .....----------
NTBGa-6(2,1) .....----------
NTBGa-7(1,1) .....----------
NTBGa-8(1,1) .....----------
NTBGa-10(1,1) .....T----------
NTBGa-11(1,1) .....----------
NTBGa-12(1,1) .....----------
NTBGa-13(1,1) .....----------
NTBGa-15(1,1) .....GTAAGTCTCCTTCCCTAAAGCGAGGGAATTCAGGGTGTCCCCATGGCATCAGCCGTGGAAATTAGTAGCTGTCCTCTCT
NTBGa-16(1,B) .....----------
NTBGb(4,1) .....----------
NTBGd(1,1) .....----------
NTBGc(2,1) .....----------
NTBGe(1,1) ----------
P2aTBGa-1(10,2) .....----------
P2aTBGa-2(3,1) .....----------
P2aTBGa-3(1,1) .....----------
P2aTBGb-1(2,2) .....----------
P2aTBGb-2(2,1) .....----------
P2aTBGb-3(1,1) .....----------
P2aTBGb-4(1,1) .....----------
P2aTBGc-1(1,B) ..G.....CAC..----------
P2aTBGc-2(1,1) ..G.....CAC..----------
P2aTBGc-3(1,1) ..G.....CAC..----------
P2aTBGc-4(1,1) ..G.....CAC..----------
P2aTBGc-6(1,1) ..G.....CAC..----------
P2aTBGc-5(1,B) ..G.....CAC..----------
15iTBGa-1(6,B) .....----------
15iTBGa-2(4,2) .....----------
15iTBGa-3(3,2) .....----------
15iTBGa-4(3,B) .....----------
15iTBGa-5(2,B) .....----------
15iTBGa-6(2,1) .....----------
15iTBGa-7(1,B) .....----------
15iTBGa-8(1,1) .....----------
15iTBGa-9(1,1) .....----------
15iTBGb-1(4,2) ----------
15iTBGb-2(2,2) ----------
15iTBGb-4(1,1) ----------
15iTBGb-5(1,B) ----------
15iTBGb-7(1,1) ----------
15iTBGc-1(3,1) ----------
6TBGa-1(10,1) .A.....TTG....CT.G---GTATGGGAGCA----------GCCATGGGATGAGAAGGTGTTCCCTCT
6TBGa-2(9,1) .A.....-...TTG....CT.G---GTATGGGAGCA----------GCCATGGGATGAGAAGGTGTTCCCTCT
6TBGa-3(3,1) .A.....TTG....CT.G---GTATGGGAGCA----------GCCATGGGATGAGAAGGTGTTCCCTCT
6TBGa-4(2,1) .AC.....CTGT..A----------GAGATGCGTGAGTCTCCCCTC-
6TBGa-5(1,1) .A.....TTG....CT.G---GTATGGGAGCA----------GCCATGGGATGAGAAGGTGTTCCCTCT
6TBGb-1(7,B) ----------
6TBGb-2(6,1) ----------
6TBGb-3(2,1) .ACA.....TTG.G..CTAGGTA----------
6TBGc-1(6,B) .....----------
6TBGc-2(3,1) .....----------
6TBGd(10,B) ----------
6TBGe(2,1) ----------

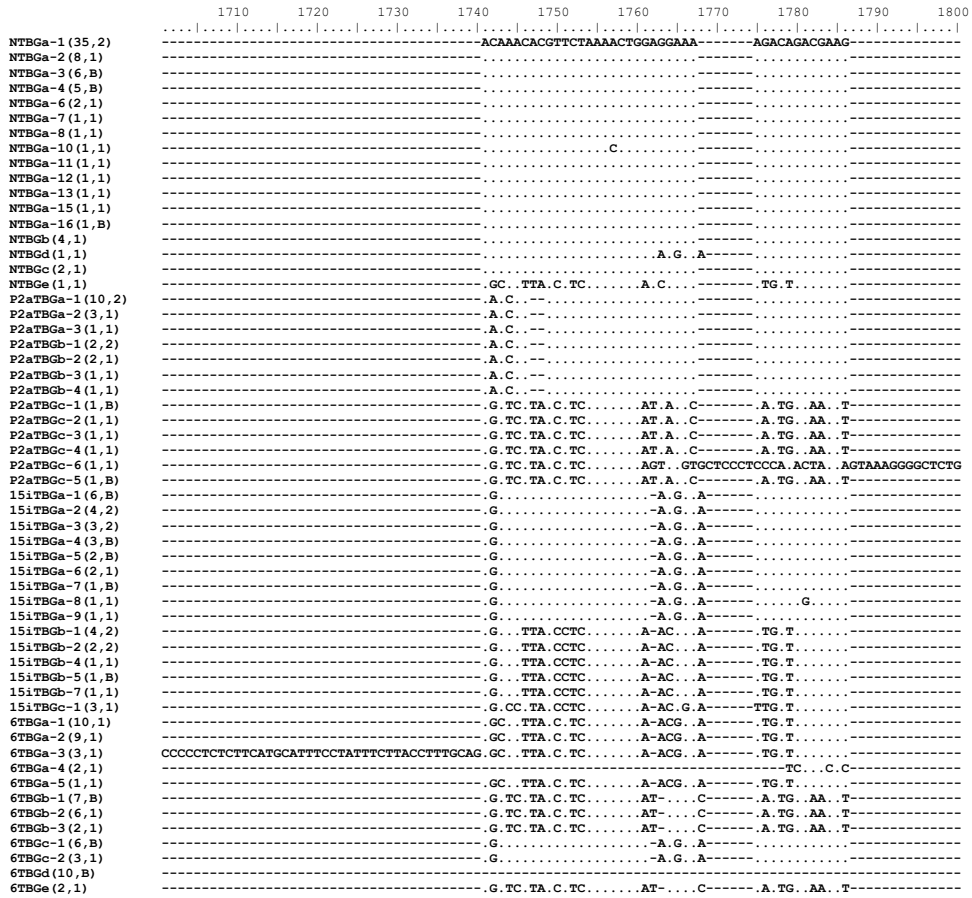

1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

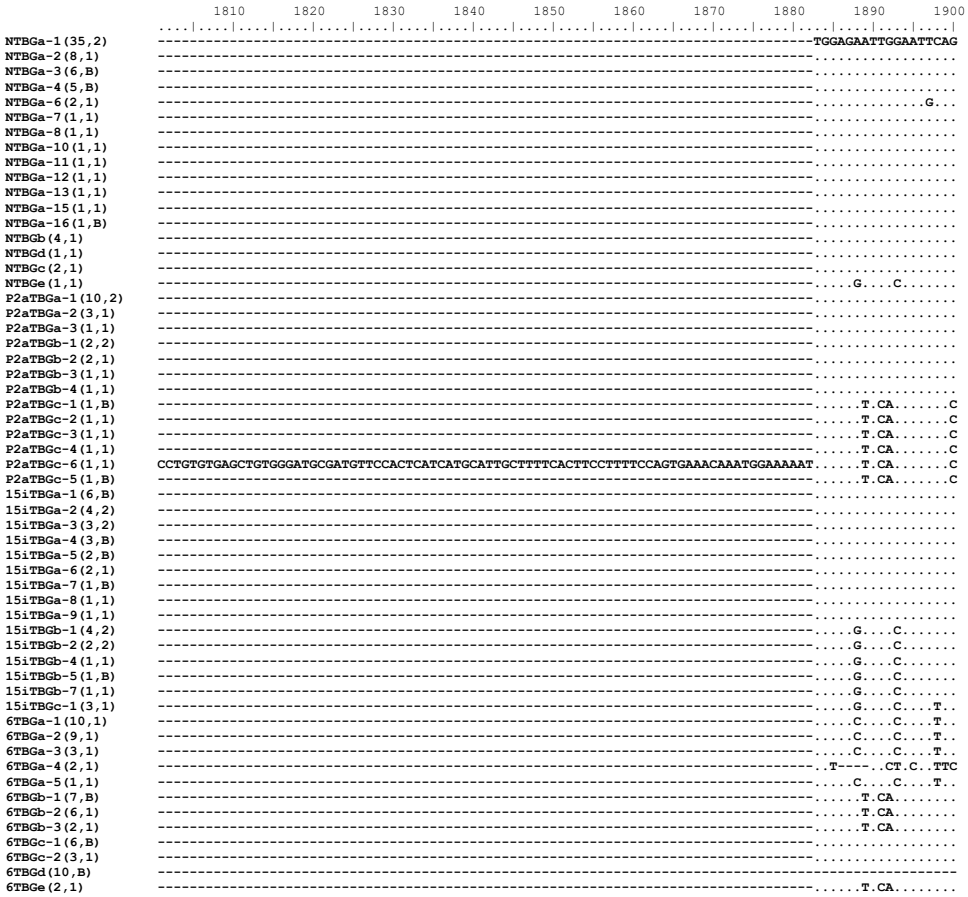

1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

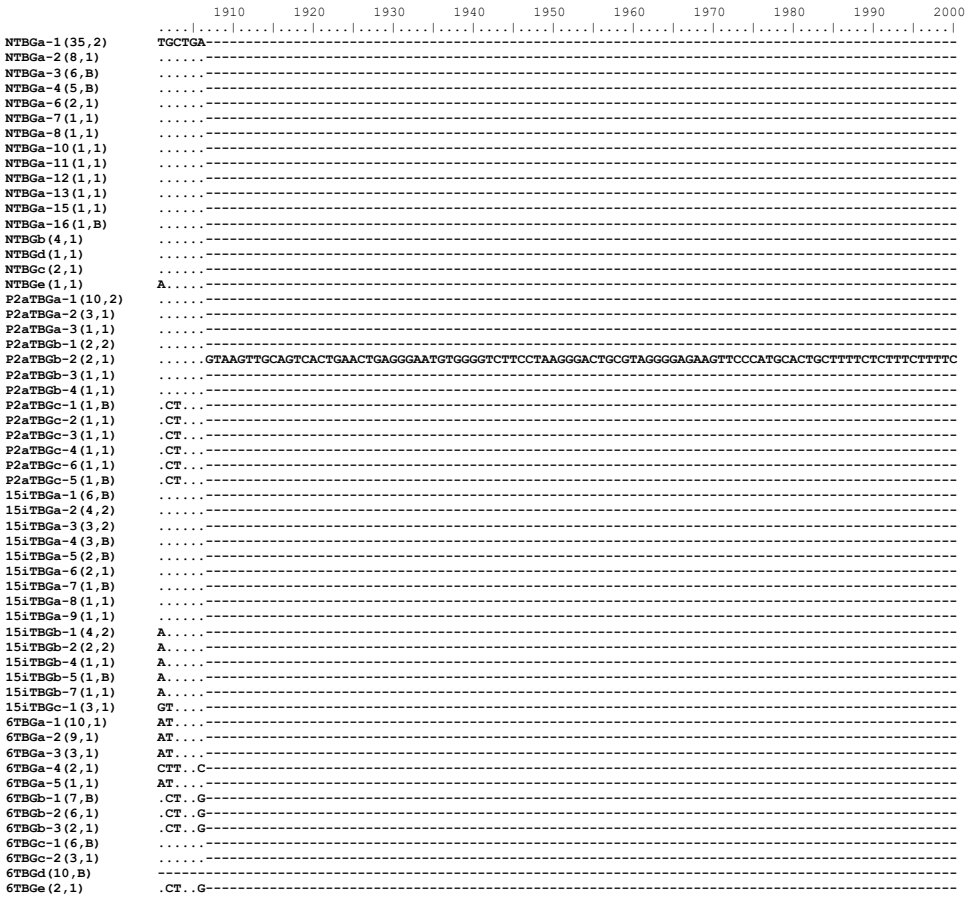

1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

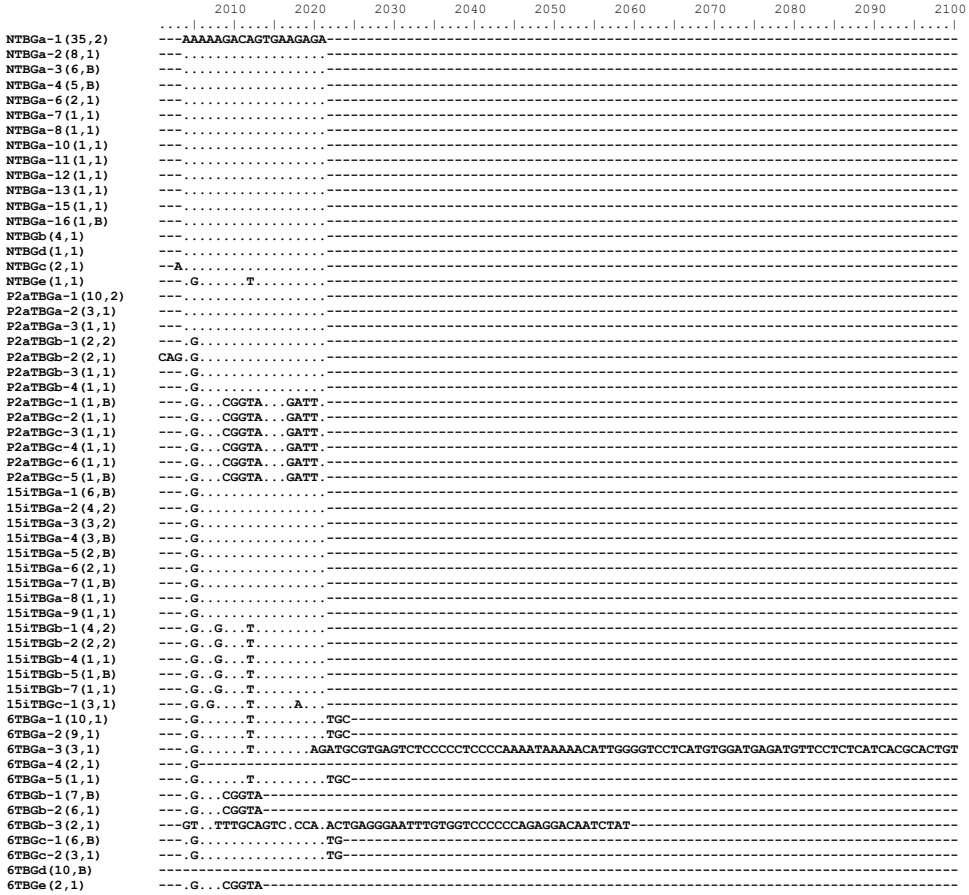

1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

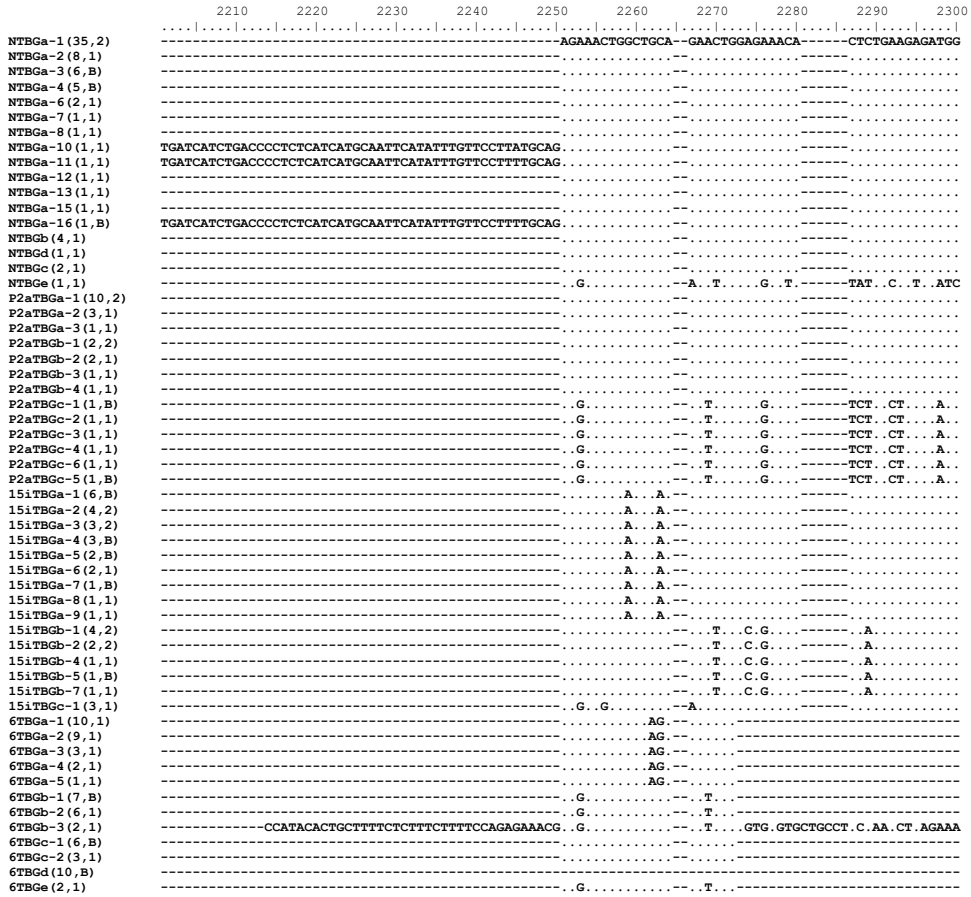

1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

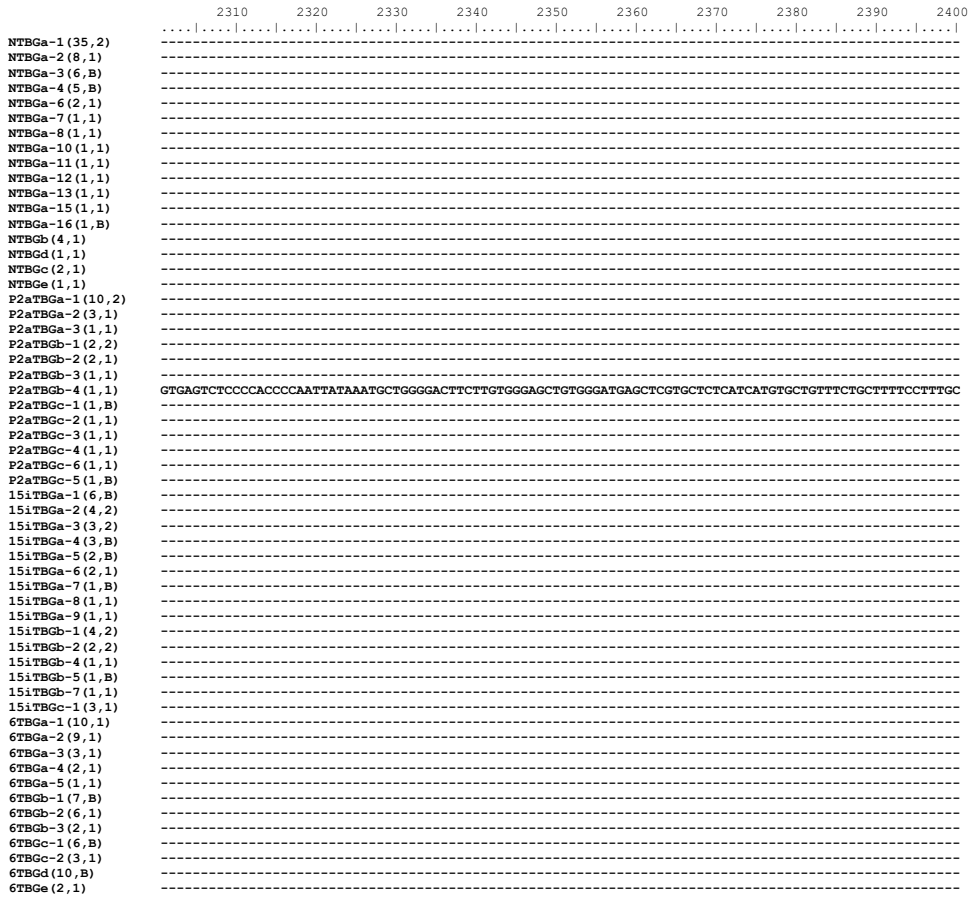

1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

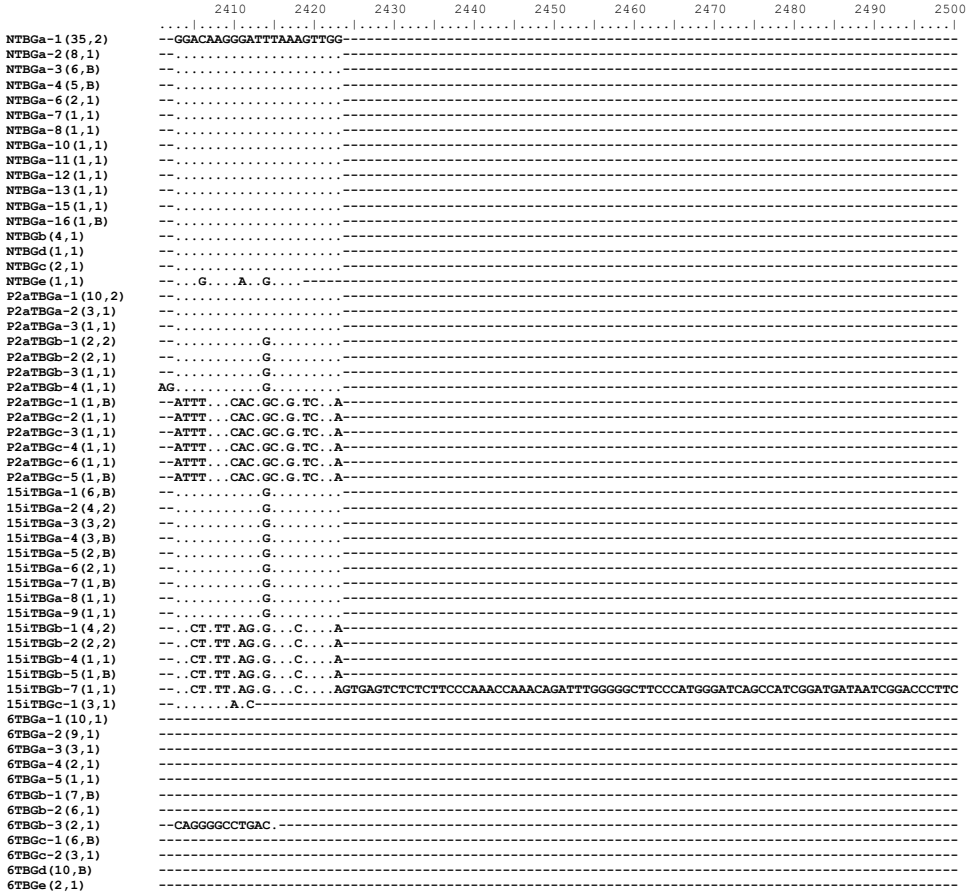

1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
NTBGa-1(35,2) TGGAGAATTGGAATTCAG
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1) G
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1) G C
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B) T CA C
P2aTBGc-2(1,1) T CA C
P2aTBGc-3(1,1) T CA C
P2aTBGc-4(1,1) T CA C
P2aTBGc-6(1,1) CCTGTGTGAGCTGTGGGATGCGATGTTCCACTCATCATGCATTGCTTTTCACTTCCTTTTCCAGTGAAACAAATGGAAAAAT T CA C
P2aTBGc-5(1,B) T CA C
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2) G C
15iTBGb-2(2,2) G C
15iTBGb-4(1,1) G C
15iTBGb-5(1,B) G C
15iTBGb-7(1,1) G C
15iTBGc-1(3,1) G C T
6TBGa-1(10,1) C C T
6TBGa-2(9,1) C C T
6TBGa-3(3,1) C C T
6TBGa-4(2,1) T CT C TTC
6TBGa-5(1,1) C C T
6TBGb-1(7,B) T CA
6TBGb-2(6,1) T CA
6TBGb-3(2,1) T CA
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1) T CA

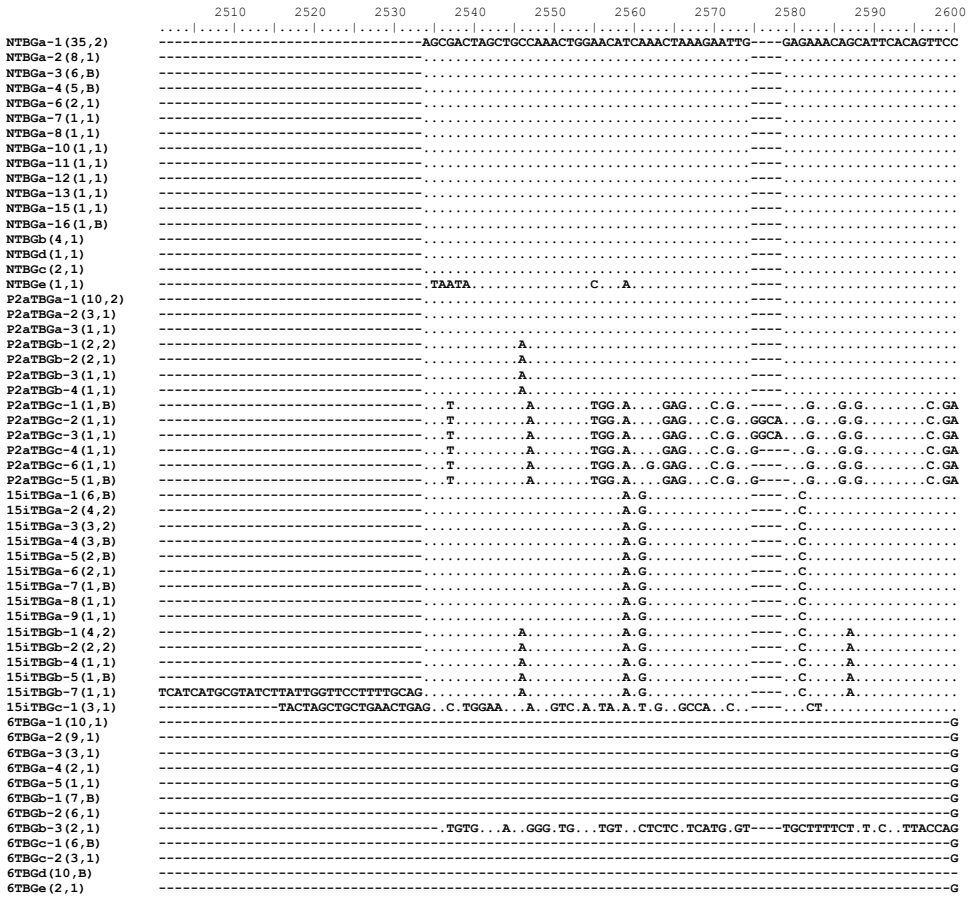

1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
NTBGa-1(35,2) TGCTGA
P2aTBGb-2(2,1) GTAAGTTGCAGTCACTGAACTGAGGGAATGTGGGGTCTTCCTAAGGGACTGCGTAGGGGAGAAGTTCCCATGCACTGCTTTTCTCTTTCTTTTC
NTBGe(1,1) A
P2aTBGc-1(1,B) CT
P2aTBGc-2(1,1) CT
P2aTBGc-3(1,1) CT
P2aTBGc-4(1,1) CT
P2aTBGc-6(1,1) CT
P2aTBGc-5(1,B) CT
15iTBGb-1(4,2) A
15iTBGb-2(2,2) A
15iTBGb-4(1,1) A
15iTBGb-5(1,B) A
15iTBGb-7(1,1) A
15iTBGc-1(3,1) GT
6TBGa-1(10,1) AT
6TBGa-2(9,1) AT
6TBGa-3(3,1) AT
6TBGa-4(2,1) CTT C
6TBGa-5(1,1) AT
6TBGb-1(7,B) CT G
6TBGb-2(6,1) CT G
6TBGb-3(2,1) CT G
6TBGe(2,1) CT G

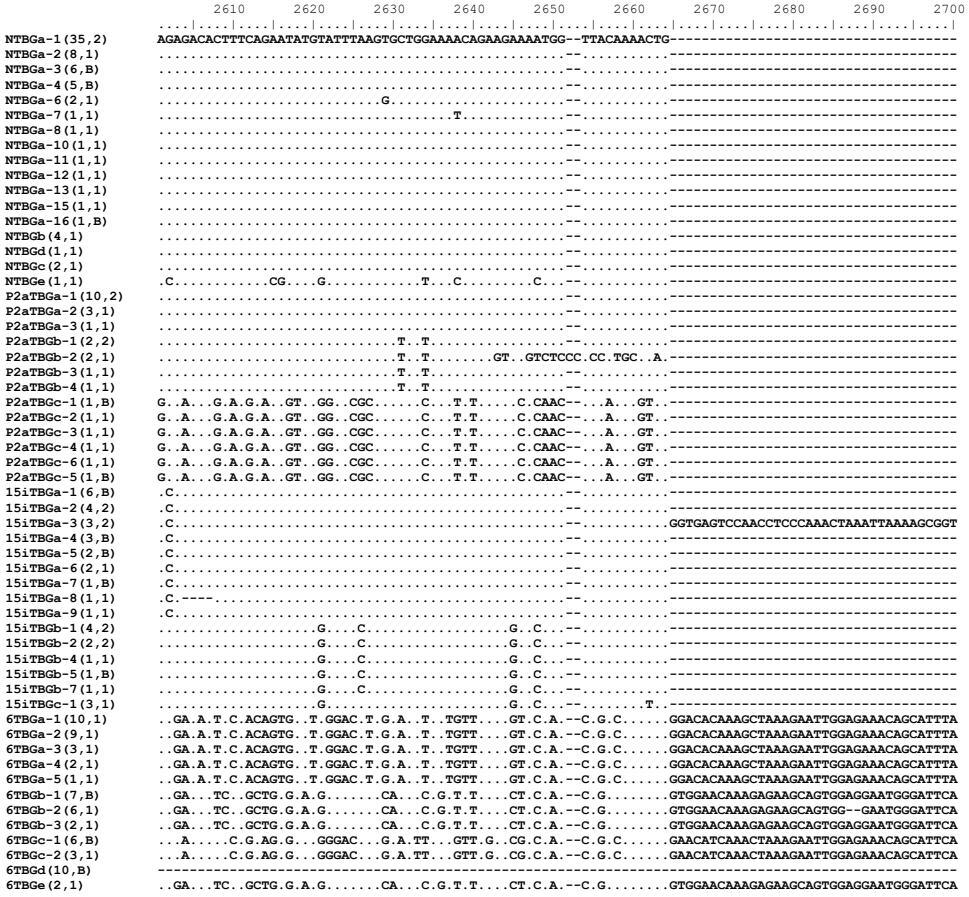

2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
NTBGa-1(35,2) ---AAAAAGACAGTGAAGAGA----------
NTBGa-2(8,1) ---..........----------
NTBGa-3(6,B) ---..........----------
NTBGa-4(5,B) ---..........----------
NTBGa-6(2,1) ---..........----------
NTBGa-7(1,1) ---..........----------
NTBGa-8(1,1) ---..........----------
NTBGa-10(1,1) ---..........----------
NTBGa-11(1,1) ---..........----------
NTBGa-12(1,1) ---..........----------
NTBGa-13(1,1) ---..........----------
NTBGa-15(1,1) ---..........----------
NTBGa-16(1,B) ---..........----------
NTBGb(4,1) ---..........----------
NTBGd(1,1) ---..........----------
NTBGc(2,1) --A..........----------
NTBGe(1,1) ---.G.....T.....----------
P2aTBGa-1(10,2) ---..........----------
P2aTBGa-2(3,1) ---..........----------
P2aTBGa-3(1,1) ---..........----------
P2aTBGb-1(2,2) ---.G.....----------
P2aTBGb-2(2,1) CAG.G.....----------
P2aTBGb-3(1,1) ---.G.....----------
P2aTBGb-4(1,1) ---.G.....----------
P2aTBGc-1(1,B) ---.G...CGGTA...GATT.----------
P2aTBGc-2(1,1) ---.G...CGGTA...GATT.----------
P2aTBGc-3(1,1) ---.G...CGGTA...GATT.----------
P2aTBGc-4(1,1) ---.G...CGGTA...GATT.----------
P2aTBGc-6(1,1) ---.G...CGGTA...GATT.----------
P2aTBGc-5(1,B) ---.G...CGGTA...GATT.----------
15iTBGa-1(6,B) ---.G.....----------
15iTBGa-2(4,2) ---.G.....----------
15iTBGa-3(3,2) ---.G.....----------
15iTBGa-4(3,B) ---.G.....----------
15iTBGa-5(2,B) ---.G.....----------
15iTBGa-6(2,1) ---.G.....----------
15iTBGa-7(1,B) ---.G.....----------
15iTBGa-8(1,1) ---.G.....----------
15iTBGa-9(1,1) ---.G.....----------
15iTBGb-1(4,2) ---.G..G...T.....----------
15iTBGb-2(2,2) ---.G..G...T.....----------
15iTBGb-4(1,1) ---.G..G...T.....----------
15iTBGb-5(1,B) ---.G..G...T.....----------
15iTBGb-7(1,1) ---.G..G...T.....----------
15iTBGc-1(3,1) ---.G.G....T....A...----------
6TBGa-1(10,1) ---.G.....T.....TGC----------
6TBGa-2(9,1) ---.G.....T.....TGC----------
6TBGa-3(3,1) ---.G.....T.....AGATGCGTGAGTCTCCCCCTCCCCAAAATAAAAACATTGGGGTCCTCATGTGGATGAGATGTTCCTCTCATCACGCACTGT
6TBGa-4(2,1) ---.G----------
6TBGa-5(1,1) ---.G.....T.....TGC----------
6TBGb-1(7,B) ---.G...CGGTA----------
6TBGb-2(6,1) ---.G...CGGTA----------
6TBGb-3(2,1) ---GT..TTTGCAGTC.CCA.ACTGAGGGAATTTGTGGTCCCCCCAGAGGACAATCTAT----------
6TBGc-1(6,B) ---.G.....TG----------
6TBGc-2(3,1) ---.G.....TG----------
6TBGd(10,B) ----------
6TBGe(2,1) ---.G...CGGTA----------
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
NTBGa-1(35,2) ----------TGGGTTATGGCTTTGGAGATCTGA----------
NTBGa-2(8,1) ----------.....----------
NTBGa-3(6,B) ----------.....----------
NTBGa-4(5,B) ----------.....----------
NTBGa-6(2,1) ----------.....----------
NTBGa-7(1,1) ----------.....----------
NTBGa-8(1,1) ----------.....----------
NTBGa-10(1,1) ----------.....GTAAGTCTCCCTCCCAACATGGAAGGAATTTATGGTCTTAGCATGGGATCAGCCATGGGA
NTBGa-11(1,1) ----------.....GTAAGTCTCCCTCCCAACATGGAAGGAATTTATGGTCTTAGCATGGGATCAGCCATGGGA
NTBGa-12(1,1) ----------.....----------
NTBGa-13(1,1) ----------.....----------
NTBGa-15(1,1) ----------.....----------
NTBGa-16(1,B) ----------.....GTAAGTCTCCCTCCCAACATGGAAGGAATTTATGGTCTTAGCATGGGATCAGCCATGGGA
NTBGb(4,1) ----------.....----------
NTBGd(1,1) ----------.....----------
NTBGc(2,1) ----------.....----------
NTBGe(1,1) ----------.....C...G...C.....----------
P2aTBGa-1(10,2) ----------.....----------
P2aTBGa-2(3,1) ----------.....----------
P2aTBGa-3(1,1) ----------.....----------
P2aTBGb-1(2,2) ----------.....C...G...C.....----------
P2aTBGb-2(2,1) ----------.....C...G...C.....----------
P2aTBGb-3(1,1) ----------.....C...G...C.....----------
P2aTBGb-4(1,1) ----------.....C...G...C.....----------
P2aTBGc-1(1,B) ----------C.G----------
P2aTBGc-2(1,1) ----------C.G----------
P2aTBGc-3(1,1) ----------C.G----------
P2aTBGc-4(1,1) ----------C.G----------
P2aTBGc-6(1,1) ----------C.G----------
P2aTBGc-5(1,B) ----------C.G----------
15iTBGa-1(6,B) ----------.....C.....C...A....----------
15iTBGa-2(4,2) ----------.....C.....C...A....----------
15iTBGa-3(3,2) ----------.....C.....C...A....----------
15iTBGa-4(3,B) ----------.....C.....C...A....----------
15iTBGa-5(2,B) ----------.....C.....C...A....----------
15iTBGa-6(2,1) ----------.....C.....C...A....----------
15iTBGa-7(1,B) ----------.....C.....C...A....----------
15iTBGa-8(1,1) ----------.....C.....C...A....----------
15iTBGa-9(1,1) ----------.....C.....C...A....----------
15iTBGb-1(4,2) ----------.....C..T.G...C...C....----------
15iTBGb-2(2,2) ----------.....C..T.G...C...C....----------
15iTBGb-4(1,1) ----------.....C..T.G...C...C....----------
15iTBGb-5(1,B) ----------.....C..T.G...C...C....----------
15iTBGb-7(1,1) ----------.....C..T.G...C...C....----------
15iTBGc-1(3,1) ----------..C...C....G...C.....----------
6TBGa-1(10,1) ----------...C....G...C.....----------
6TBGa-2(9,1) ----------...C....G...C.....----------
6TBGa-3(3,1) TTCTACTTTTCCTTTGCA....C....G...C.....----------
6TBGa-4(2,1) ----------...C....G...C.....----------
6TBGa-5(1,1) ----------...C....G...C.....----------
6TBGb-1(7,B) ----------..AGA..AC.G----------
6TBGb-2(6,1) ----------..AGA..AC.G----------
6TBGb-3(2,1) ----------.GG.CAAAAAATCCCC.....----------
6TBGc-1(6,B) ----------.....C...A....----------
6TBGc-2(3,1) ----------.....C...A....----------
6TBGd(10,B) ----------
6TBGe(2,1) ----------..AGA..AC.G----------

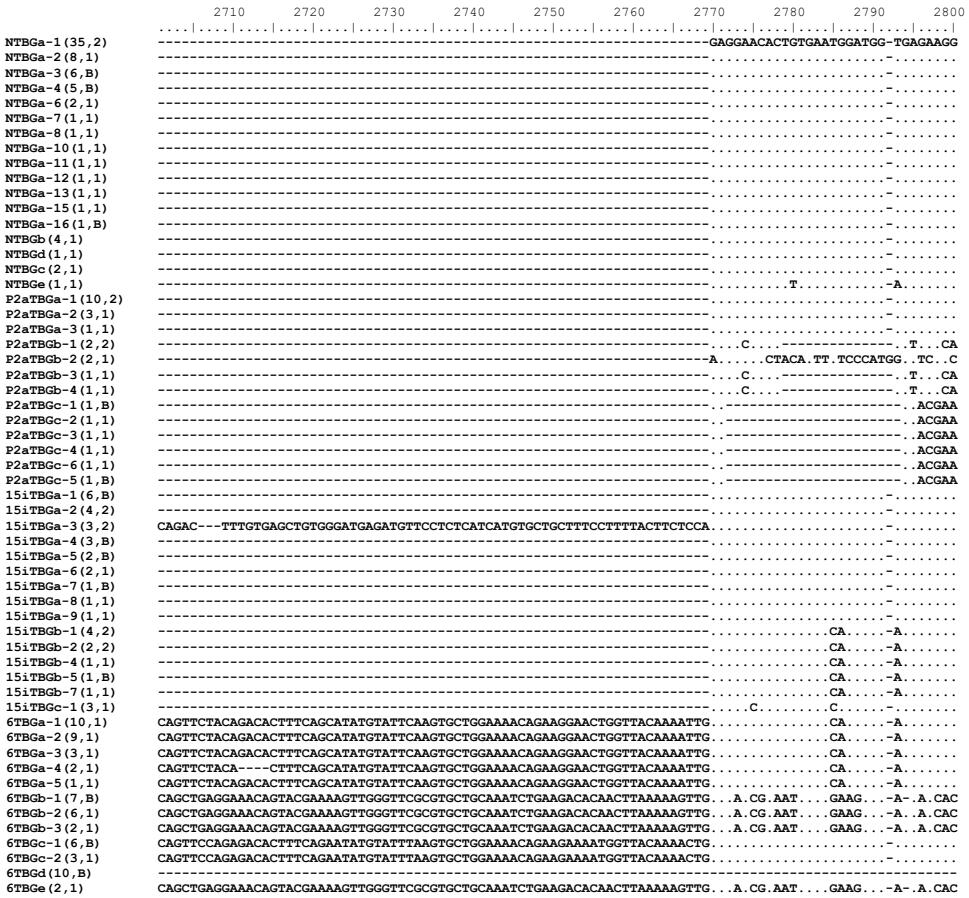

2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

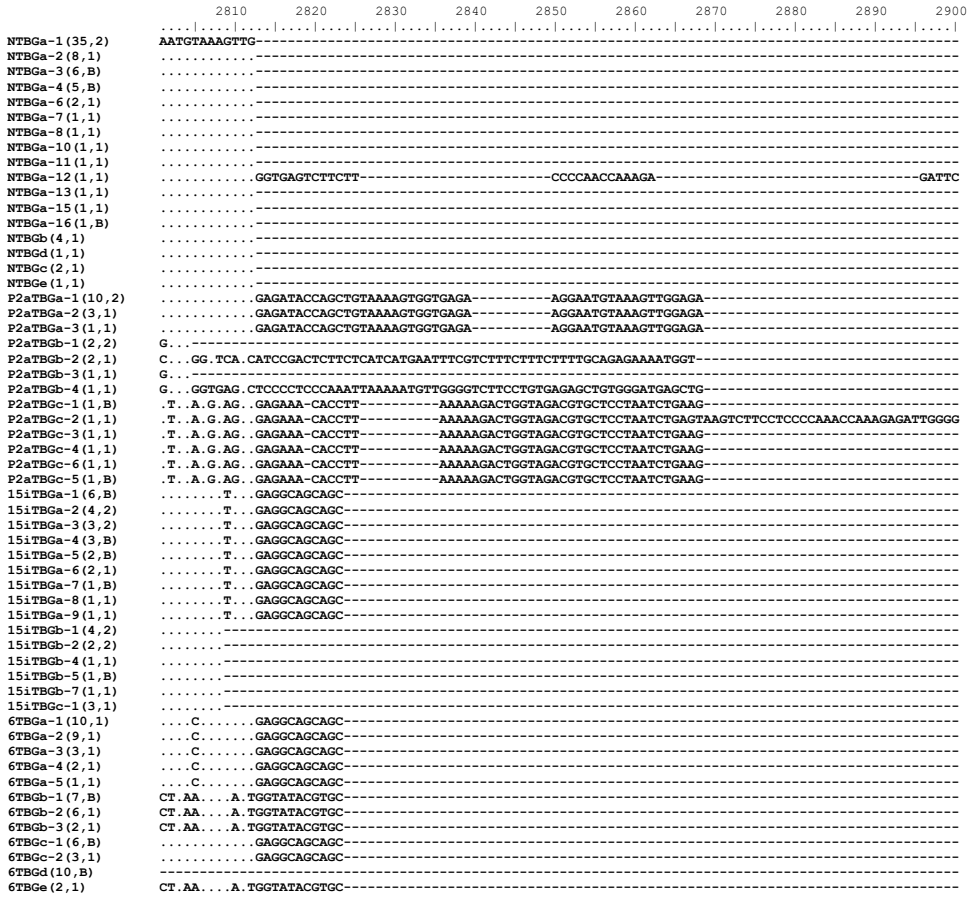

2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

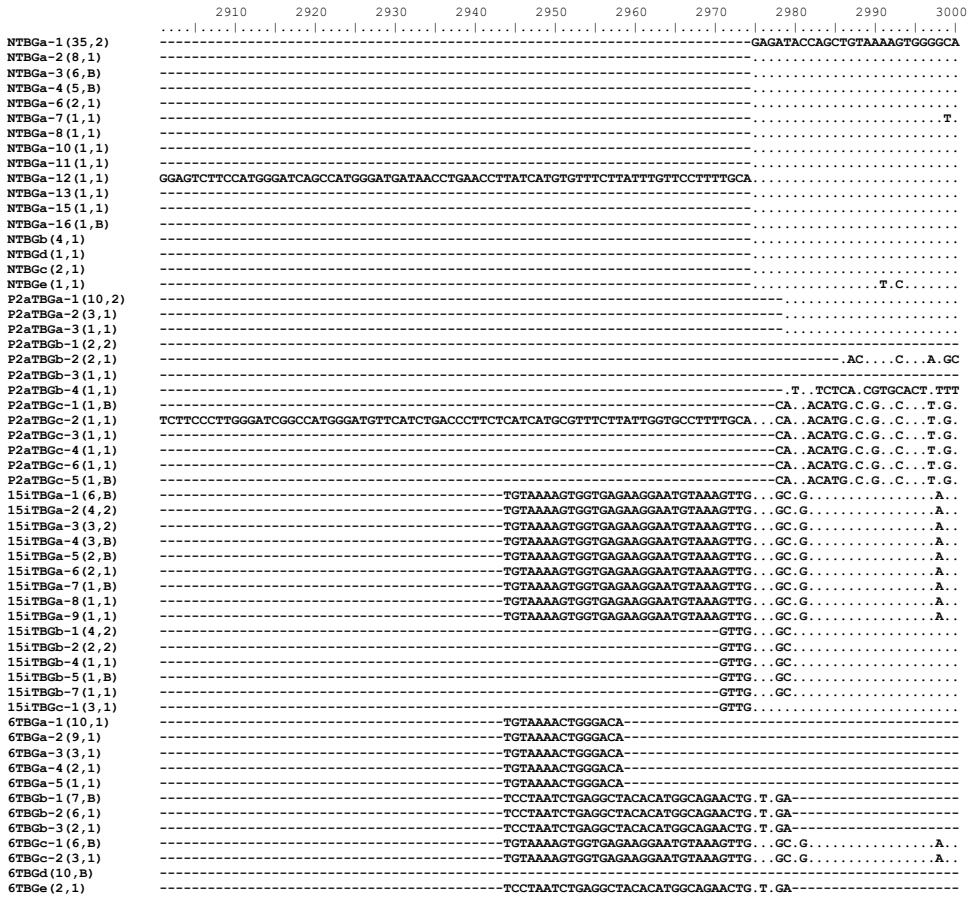

2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

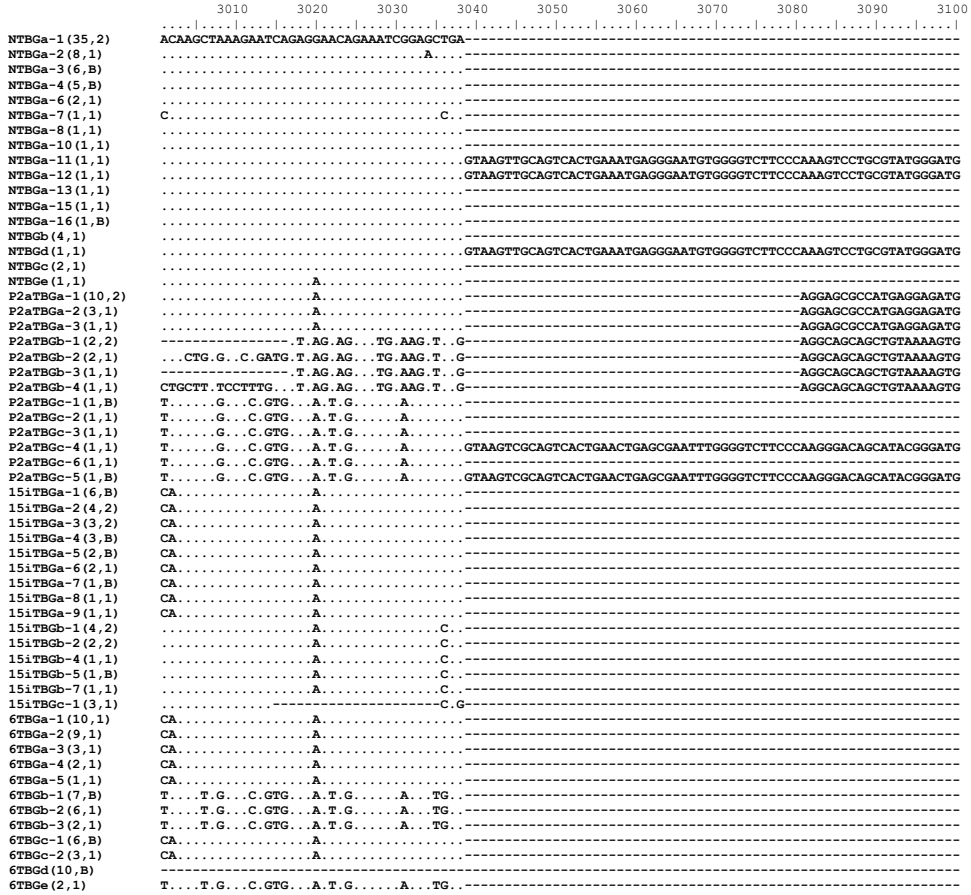

2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

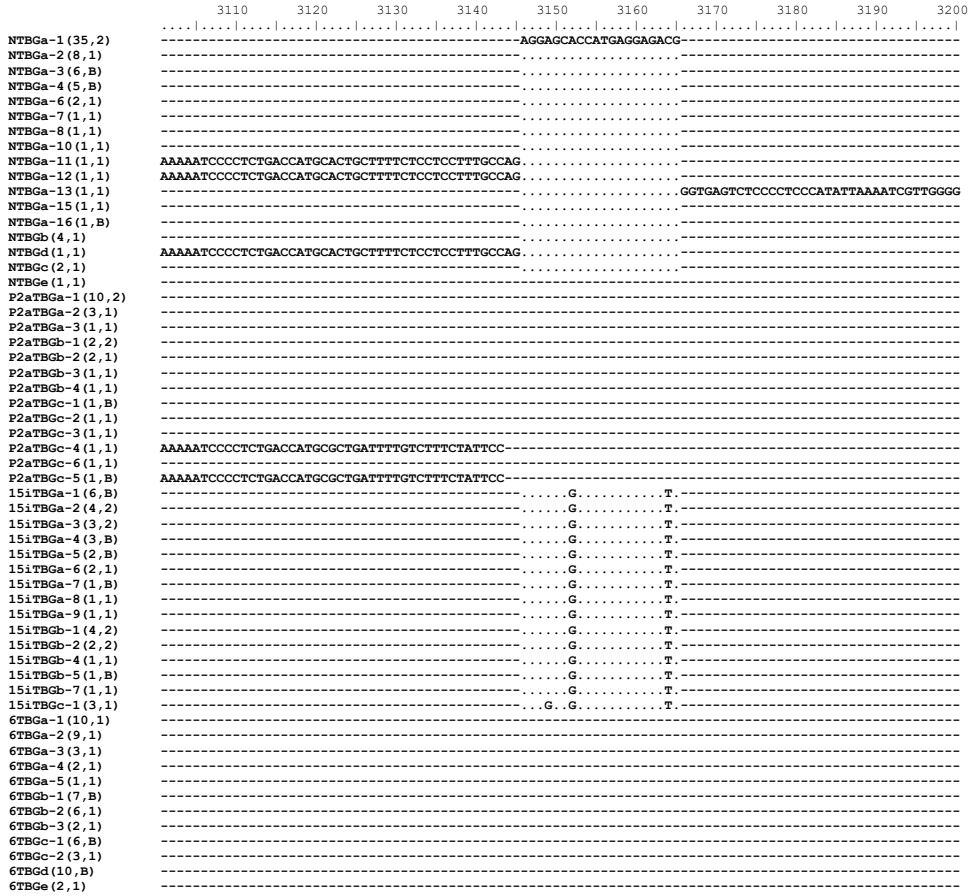

2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
NTBGa-1(35,2) AGAGACACTTTCAGAATATGTATTTAAGTGCTGGAAAACAGAAGAAAATGG--TTACAAAACTG
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2) GGTGAGTCCAACCTCCCAAACTAAATTAAAAGCGGT
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1) GGACACAAAGCTAAAGAATTGGAGAAACAGCATTTA
6TBGa-2(9,1) GGACACAAAGCTAAAGAATTGGAGAAACAGCATTTA
6TBGa-3(3,1) GGACACAAAGCTAAAGAATTGGAGAAACAGCATTTA
6TBGa-4(2,1) GGACACAAAGCTAAAGAATTGGAGAAACAGCATTTA
6TBGa-5(1,1) GGACACAAAGCTAAAGAATTGGAGAAACAGCATTTA
6TBGb-1(7,B) GTGGAACAAAGAGAAGCAGTGGAGGAATGGGATTCA
6TBGb-2(6,1) GTGGAACAAAGAGAAGCAGTGG--GAATGGGATTCA
6TBGb-3(2,1) GTGGAACAAAGAGAAGCAGTGGAGGAATGGGATTCA
6TBGc-1(6,B) GAACATCAAACTAAAGAATTGGAGAAACAGCATTCA
6TBGc-2(3,1) GAACATCAAACTAAAGAATTGGAGAAACAGCATTCA
6TBGd(10,B)
6TBGe(2,1) GTGGAACAAAGAGAAGCAGTGGAGGAATGGGATTCA

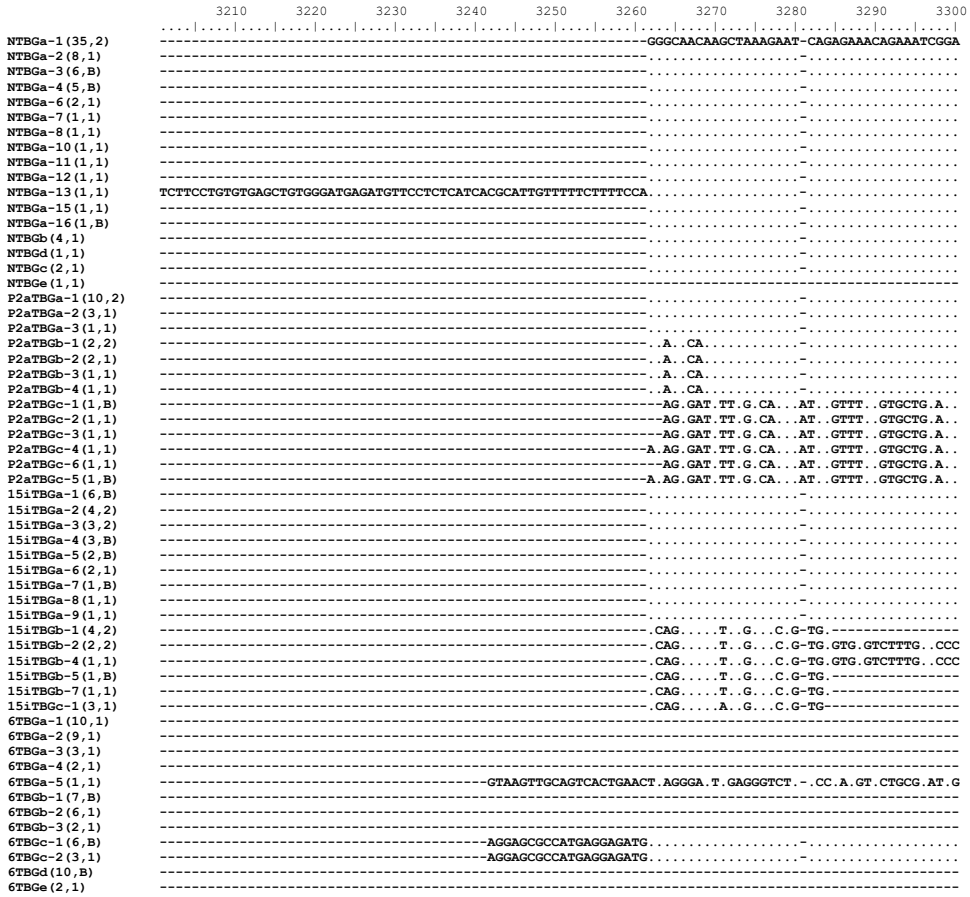

2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
NTBGa-1(35,2) GAGGAACACTGTGAATGGATGG-TGAGAAGG
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2) CAGAC---TTTGTGAGCTGTGGGATGAGATGTTCCTCTCATCATGTGCTGCTTTCCTTTTACTTCTCCA
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1) CAGTTCTACAGACACTTTCAGCATATGTATTCAAGTGCTGGAAAACAGAAGGAACTGGTTACAAAATTG
6TBGa-2(9,1) CAGTTCTACAGACACTTTCAGCATATGTATTCAAGTGCTGGAAAACAGAAGGAACTGGTTACAAAATTG
6TBGa-3(3,1) CAGTTCTACAGACACTTTCAGCATATGTATTCAAGTGCTGGAAAACAGAAGGAACTGGTTACAAAATTG
6TBGa-4(2,1) CAGTTCTACA----CTTTCAGCATATGTATTCAAGTGCTGGAAAACAGAAGGAACTGGTTACAAAATTG
6TBGa-5(1,1) CAGTTCTACAGACACTTTCAGCATATGTATTCAAGTGCTGGAAAACAGAAGGAACTGGTTACAAAATTG
6TBGb-1(7,B) CAGCTGAGGAAACAGTACGAAAAGTTGGGTTCGCGTGCTGCAAATCTGAAGACACAACTTAAAAAGTTG...A.CG.AAT....GAAG...-A-.A.CAC
6TBGb-2(6,1) CAGCTGAGGAAACAGTACGAAAAGTTGGGTTCGCGTGCTGCAAATCTGAAGACACAACTTAAAAAGTTG...A.CG.AAT....GAAG...-A..A.CAC
6TBGb-3(2,1) CAGCTGAGGAAACAGTACGAAAAGTTGGGTTCGCGTGCTGCAAATCTGAAGACACAACTTAAAAAGTTG...A.CG.AAT....GAAG...-A..A.CAC
6TBGc-1(6,B) CAGTTCCAGAGACACTTTCAGAATATGTATTTAAGTGCTGGAAAACAGAAGAAAATGGTTACAAAACTG
6TBGc-2(3,1) CAGTTCCAGAGACACTTTCAGAATATGTATTTAAGTGCTGGAAAACAGAAGAAAATGGTTACAAAACTG
6TBGd(10,B)
6TBGe(2,1) CAGCTGAGGAAACAGTACGAAAAGTTGGGTTCGCGTGCTGCAAATCTGAAGACACAACTTAAAAAGTTG...A.CG.AAT....GAAG...-A-.A.CAC

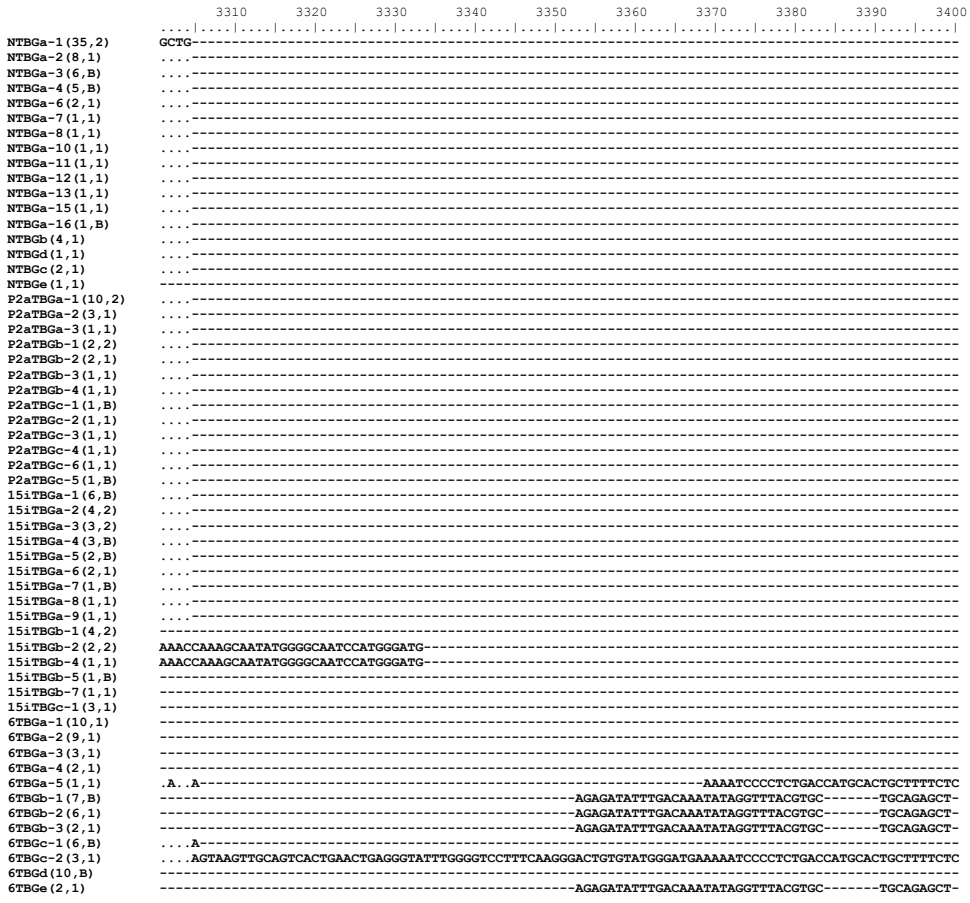

2810 2820 2830 2840 2850 2860 2870 2880 2890 2900
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

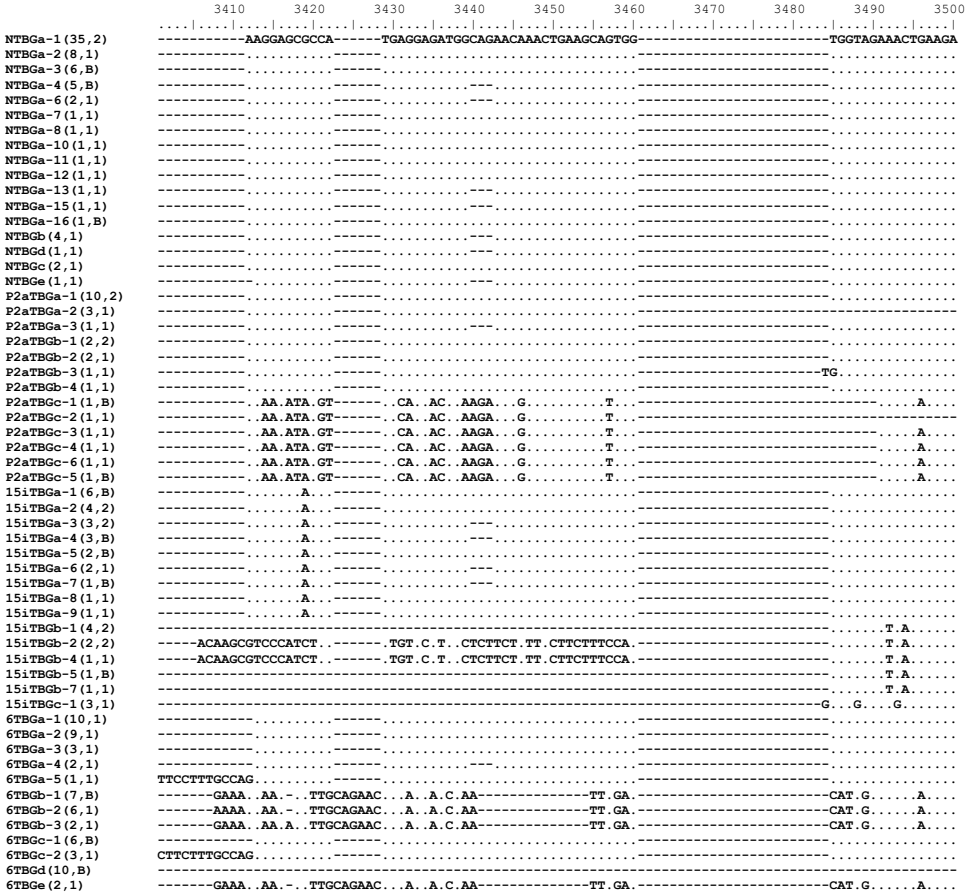

2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

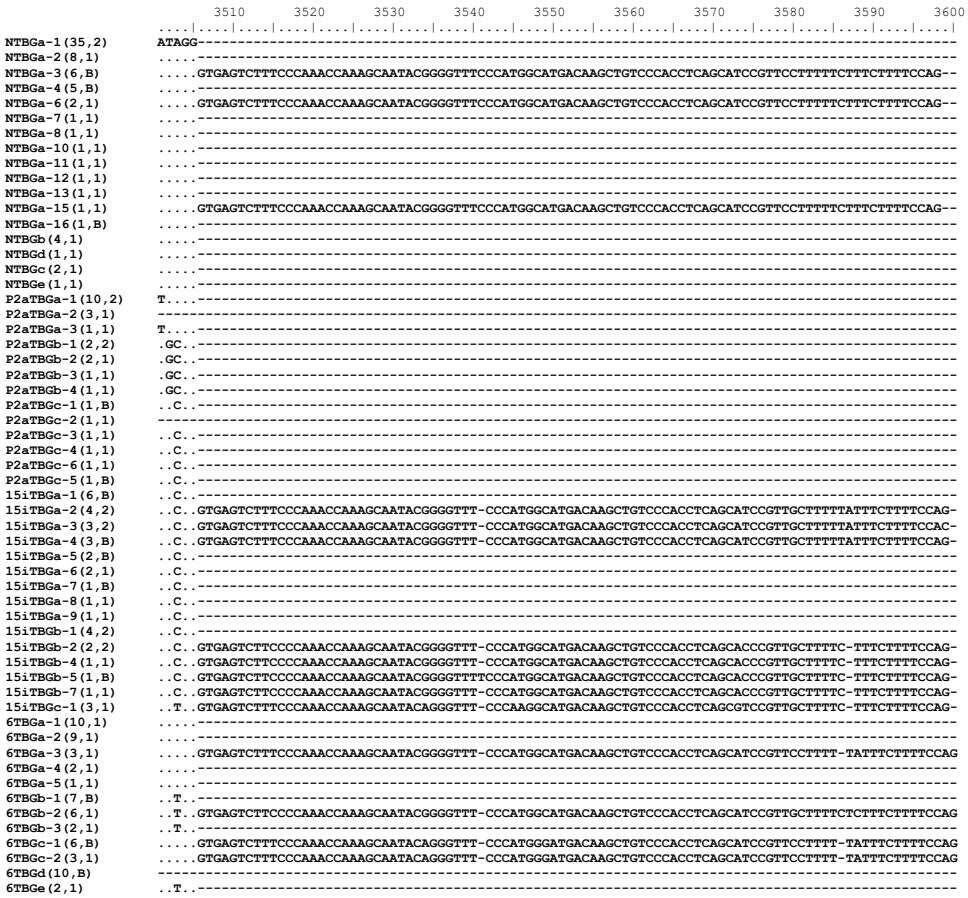

3010 3020 3030 3040 3050 3060 3070 3080 3090 3100
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

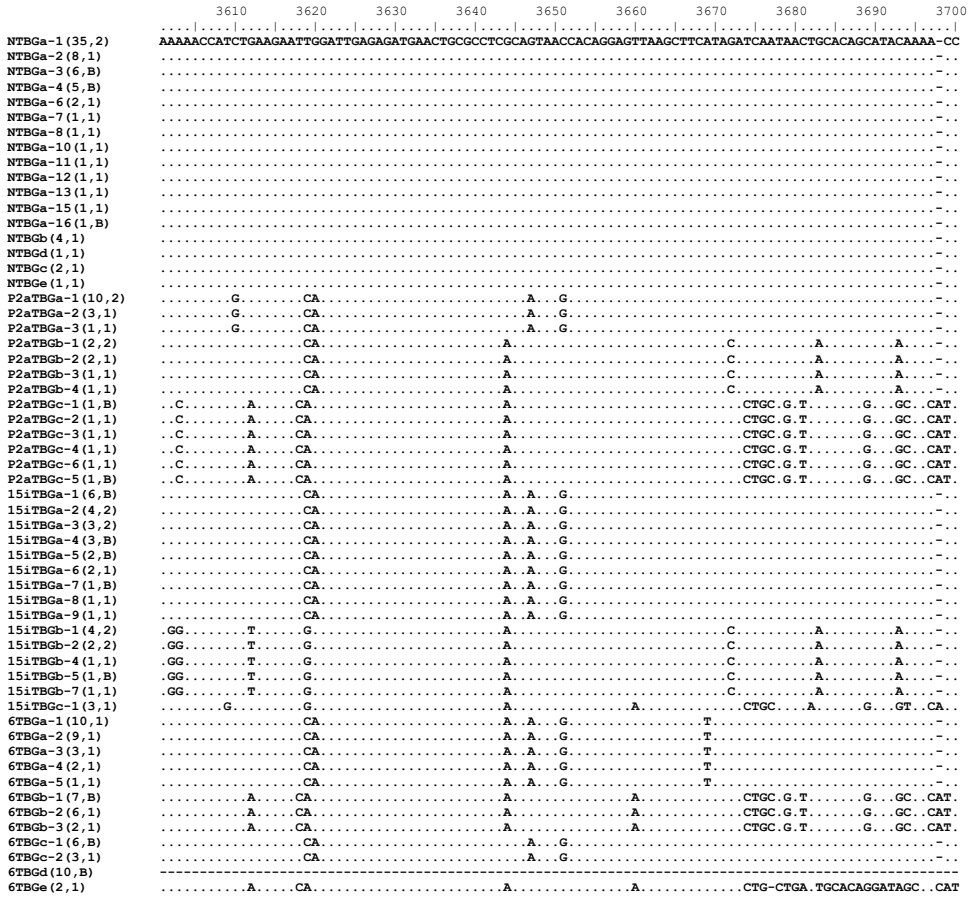

3110 3120 3130 3140 3150 3160 3170 3180 3190 3200
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

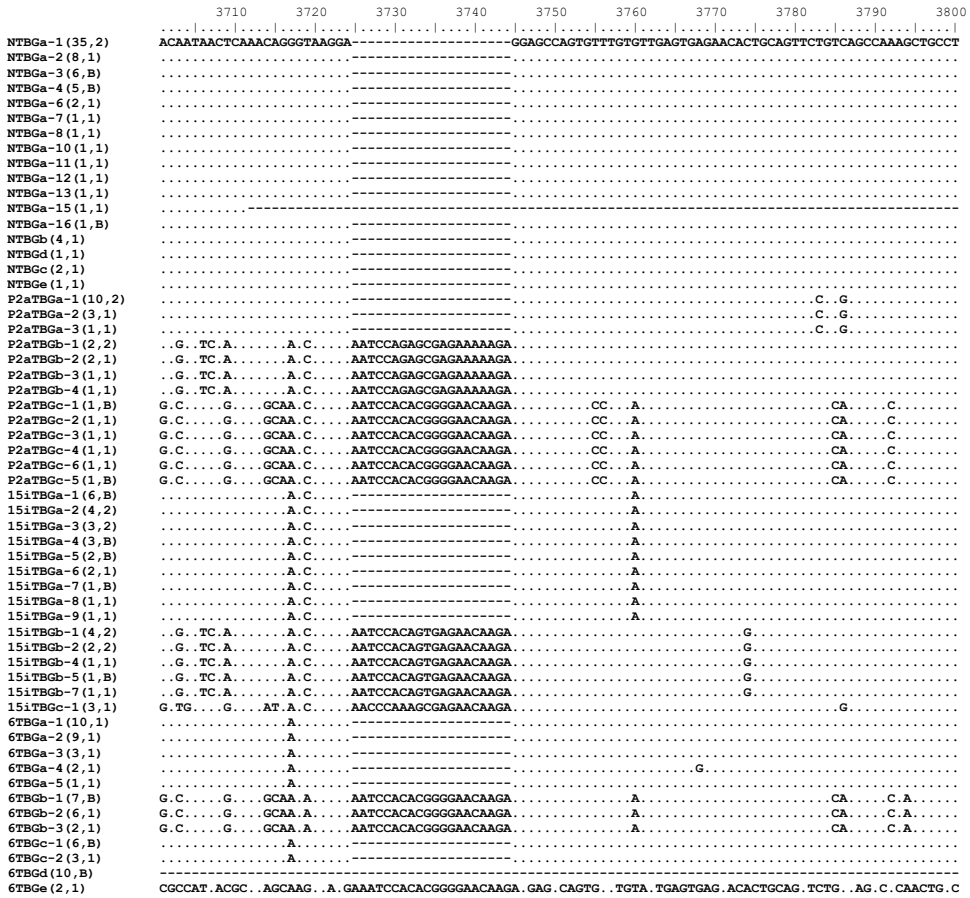

3210 3220 3230 3240 3250 3260 3270 3280 3290 3300
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)

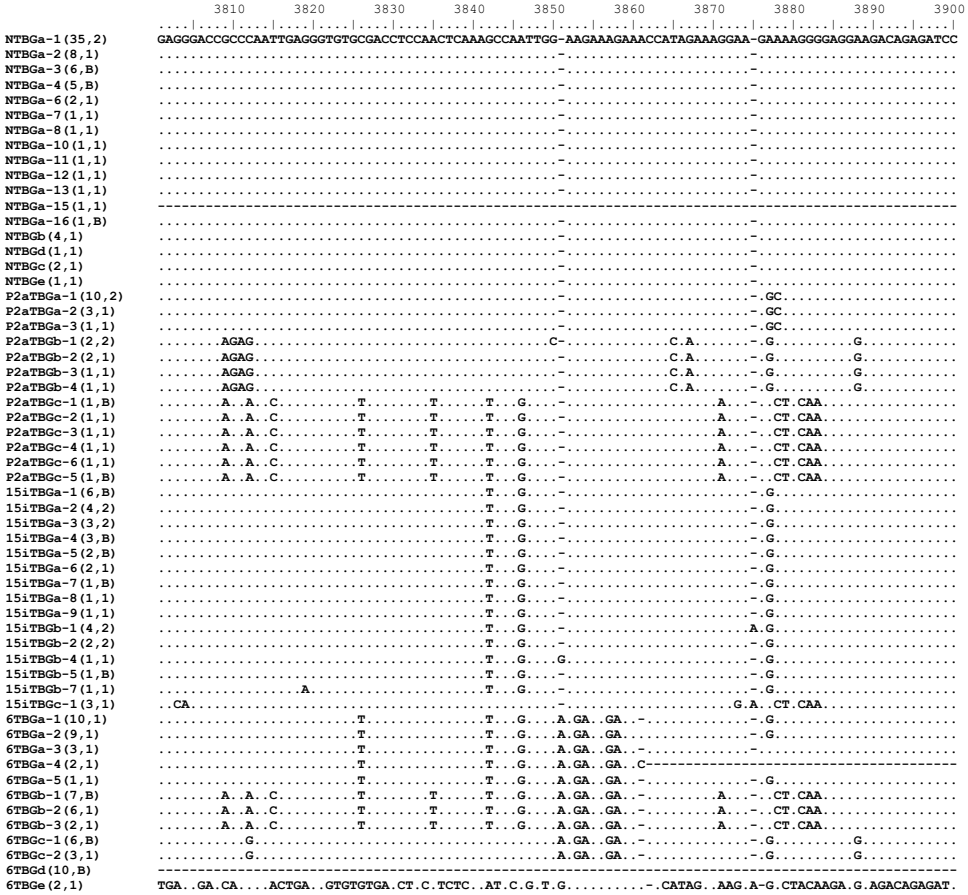

3310 3320 3330 3340 3350 3360 3370 3380 3390 3400
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)


3410 3420 3430 3440 3450 3460 3470 3480 3490 3500
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)


3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
NTBGa-1(35,2)
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1)


3610 3620 3630 3640 3650 3660 3670 3680 3690 3700
NTBGa-1(35,2) AAAAACCATCTGAAGAATTGGATTGAGAGATGAACTGCGCCTCGCAGTAACCACAGGAGTTAAGCTTCATAGATCAATAACTGCACAGCATACAAAA-CC
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1) ...........A.....CA.....A.....A.....CTG-CTGA.TGCACAGGATAGC..CAT


3710 3720 3730 3740 3750 3760 3770 3780 3790 3800
NTBGa-1(35,2) ACAATAACTCAAACAGGGTAAGGA----------GGAGCCAGTGTTTGTGTTGAGTGAGAACACTGCAGTTCTGTCAGCCAAAGCTGCCT
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1) CGCCAT.ACGC..AGCAAG..A.GAAATCCACACGGGGAACAAGA.GAG.CAGTG..TGTA.TGAGTGAG.ACACTGCAG.TCTG..AG.C.CAACTG.C


3810 3820 3830 3840 3850 3860 3870 3880 3890 3900
NTBGa-1(35,2) GAGGGACCGCCCAATTGAGGGTGTGCGACCTCCAACTCAAAGCCAATTGG-AAGAAAGAAACCATAGAAAGGAA-GAAAAGGGGAGGAAGACAGAGATCC
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1) TGA..GA.CA....ACTGA..GTGTGTGA.CT.C.TCTC..AT.C.G.T.G.....-.CATAG..AAG.A-G.CTACAAGA.G.AGACAGAGAT.
3910 3920 3930 3940 3950 3960 3970 3980 3990 4000
NTBGa-1(35,2) TGGAAGAGATATGGGCATTTGGGGAAATAGTGTGACCATGTATCAGGCTTTGTGGACATCTAACGAATATGTCATGTTTTTGTAAATACAAGCATGCACG
NTBGa-2(8,1)
NTBGa-3(6,B)
NTBGa-4(5,B)
NTBGa-6(2,1)
NTBGa-7(1,1)
NTBGa-8(1,1)
NTBGa-10(1,1)
NTBGa-11(1,1)
NTBGa-12(1,1)
NTBGa-13(1,1)
NTBGa-15(1,1)
NTBGa-16(1,B)
NTBGb(4,1)
NTBGd(1,1)
NTBGc(2,1)
NTBGe(1,1)
P2aTBGa-1(10,2)
P2aTBGa-2(3,1)
P2aTBGa-3(1,1)
P2aTBGb-1(2,2)
P2aTBGb-2(2,1)
P2aTBGb-3(1,1)
P2aTBGb-4(1,1)
P2aTBGc-1(1,B)
P2aTBGc-2(1,1)
P2aTBGc-3(1,1)
P2aTBGc-4(1,1)
P2aTBGc-6(1,1)
P2aTBGc-5(1,B)
15iTBGa-1(6,B)
15iTBGa-2(4,2)
15iTBGa-3(3,2)
15iTBGa-4(3,B)
15iTBGa-5(2,B)
15iTBGa-6(2,1)
15iTBGa-7(1,B)
15iTBGa-8(1,1)
15iTBGa-9(1,1)
15iTBGb-1(4,2)
15iTBGb-2(2,2)
15iTBGb-4(1,1)
15iTBGb-5(1,B)
15iTBGb-7(1,1)
15iTBGc-1(3,1)
6TBGa-1(10,1)
6TBGa-2(9,1)
6TBGa-3(3,1)
6TBGa-4(2,1)
6TBGa-5(1,1)
6TBGb-1(7,B)
6TBGb-2(6,1)
6TBGb-3(2,1)
6TBGc-1(6,B)
6TBGc-2(3,1)
6TBGd(10,B)
6TBGe(2,1) CT.GGA.AGGGACA.ACA..TT..G...TAACATGG.CATGTATCA.GGG..A...AT.....T........A.G.C..C..G.....T..A.......A

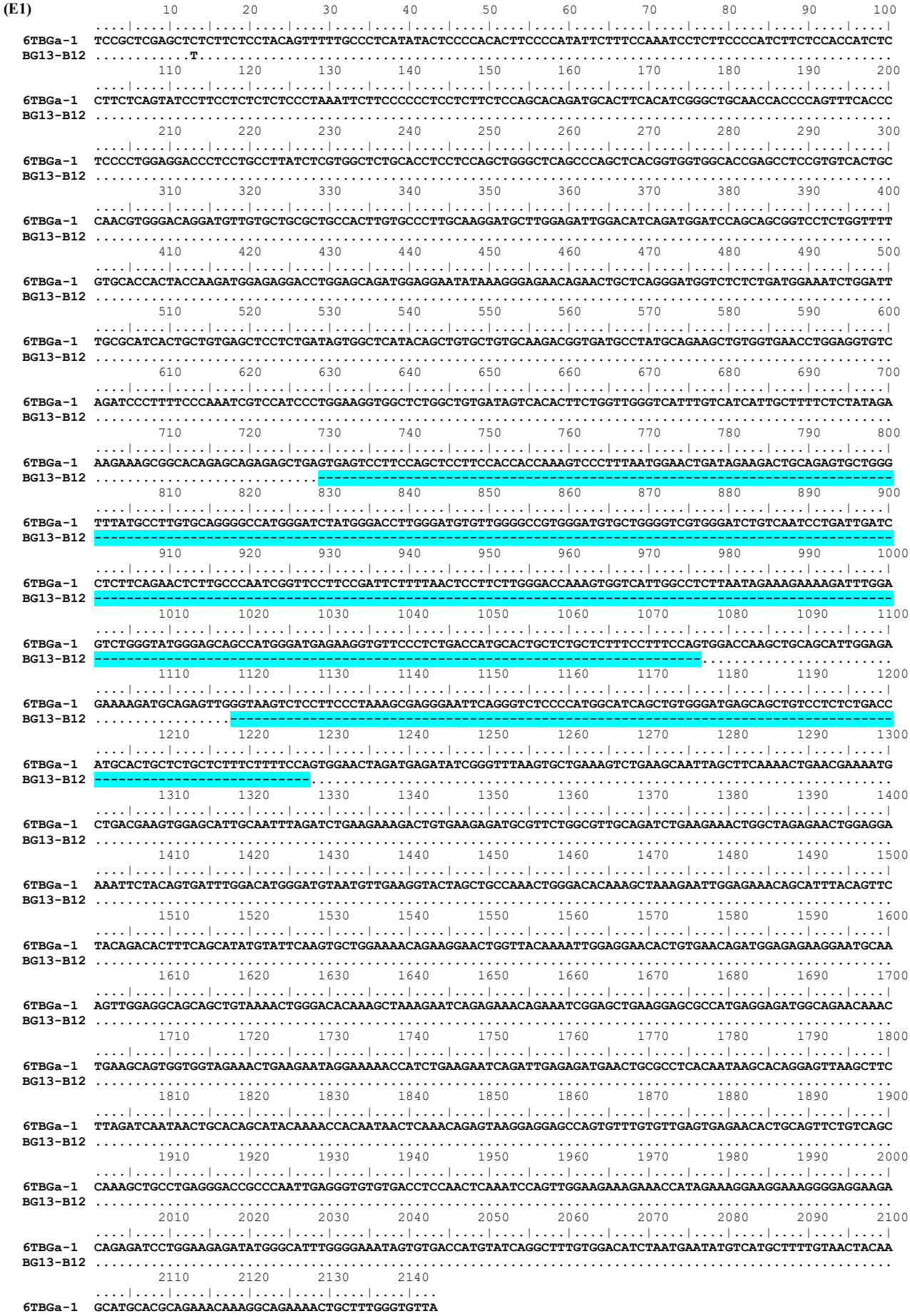
```
                                4010      4020      4030
                           ....|....|....|....|....|....|....
NTBGa-1(35,2)              CAGAAACAAAGGGAGAAAACTGCTTTGGGTGTTA
NTBGa-2(8,1)               ..................................
NTBGa-3(6,B)               ..................................
NTBGa-4(5,B)               ..................................
NTBGa-6(2,1)               ..................................
NTBGa-7(1,1)               ..................................
NTBGa-8(1,1)               ..................................AT
NTBGa-10(1,1)              ..................................
NTBGa-11(1,1)              ..................................
NTBGa-12(1,1)              ..................................
NTBGa-13(1,1)              ..................................
NTBGa-15(1,1)              ----------------------------------
NTBGa-16(1,B)              ..................................
NTBGb(4,1)                 ..................................
NTBGd(1,1)                 ..................................
NTBGc(2,1)                 ..................................
NTBGe(1,1)                 ..................................
P2aTBGa-1(10,2)            ............C.....................
P2aTBGa-2(3,1)             ............T.....................
P2aTBGa-3(1,1)             ............T.....................
P2aTBGb-1(2,2)             ............T.....................
P2aTBGb-2(2,1)             ............T.....................
P2aTBGb-3(1,1)             ............C.....................
P2aTBGb-4(1,1)             ............C.....................
P2aTBGc-1(1,B)             .....GT.....C.....................
P2aTBGc-2(1,1)             .....GT.....C.....................
P2aTBGc-3(1,1)             .....GT.....C.....................
P2aTBGc-4(1,1)             .....GT.....C.....................
P2aTBGc-6(1,1)             .....GT.....C.....................
P2aTBGc-5(1,B)             .....GT.....T.....................
15iTBGa-1(6,B)             ............C.....................
15iTBGa-2(4,2)             ............C.....................
15iTBGa-3(3,2)             ............C.....................
15iTBGa-4(3,B)             ............C.....................
15iTBGa-5(2,B)             ............C.....................
15iTBGa-6(2,1)             ............C.....................
15iTBGa-7(1,B)             ............C.....................
15iTBGa-8(1,1)             ............C.....................
15iTBGa-9(1,1)             ............T.....................
15iTBGb-1(4,2)             ............C.....................
15iTBGb-2(2,2)             ............C.....................
15iTBGb-4(1,1)             ............C.....................
15iTBGb-5(1,B)             ............T.....................
15iTBGb-7(1,1)             ............C.....................
15iTBGc-1(3,1)             ..................................
6TBGa-1(10,1)              ............C.....................
6TBGa-2(9,1)               ............T.....................
6TBGa-3(3,1)               ............T.....................
6TBGa-4(2,1)               ............C.....................
6TBGa-5(1,1)               ............T.....................
6TBGb-1(7,B)               .....GT.....T.....................
6TBGb-2(6,1)               .....GT.....T.....................
6TBGb-3(2,1)               .....GT.....C.....................
6TBGc-1(6,B)               ............C.....................
6TBGc-2(3,1)               ............T.....................
6TBGd(10,B)                ----------------------------------
6TBGe(2,1)                 .....GT.....C.....................
```

Appendix D. Alignment of nucleotide sequences from all 57 different clones considered to be confirmed (that is, with obvious chimeras not included) from T cells of four chicken lines, including both exons and apparent introns found in the clones. Names of the transcripts follow the convention: abbreviated line name, “T” for T cells, “BG”, a letter representing the exon 2 sequence with “a” being the most frequently detected exon 2 sequence (and “b” being the second most frequently detected exon 2 sequence, and so forth), a dash and then a number representing the alternative splicing variant with “1” being the most frequently detected clone (and “2” the second most frequently detected clone, and so forth). Numbers in parentheses indicate the number of clones found for a particular full sequence, followed by the number of independent PCRs in which the sequence was identified (1, found in one PCR; 2 found in 2 PCRs; B, found in one PCR described in this paper and one using B cell cDNA, data not shown). Letters indicate nucleotides, dots indicate identities with NTBGa-1 sequence, dashes indicate no sequence present compared to one or more of the other sequences.

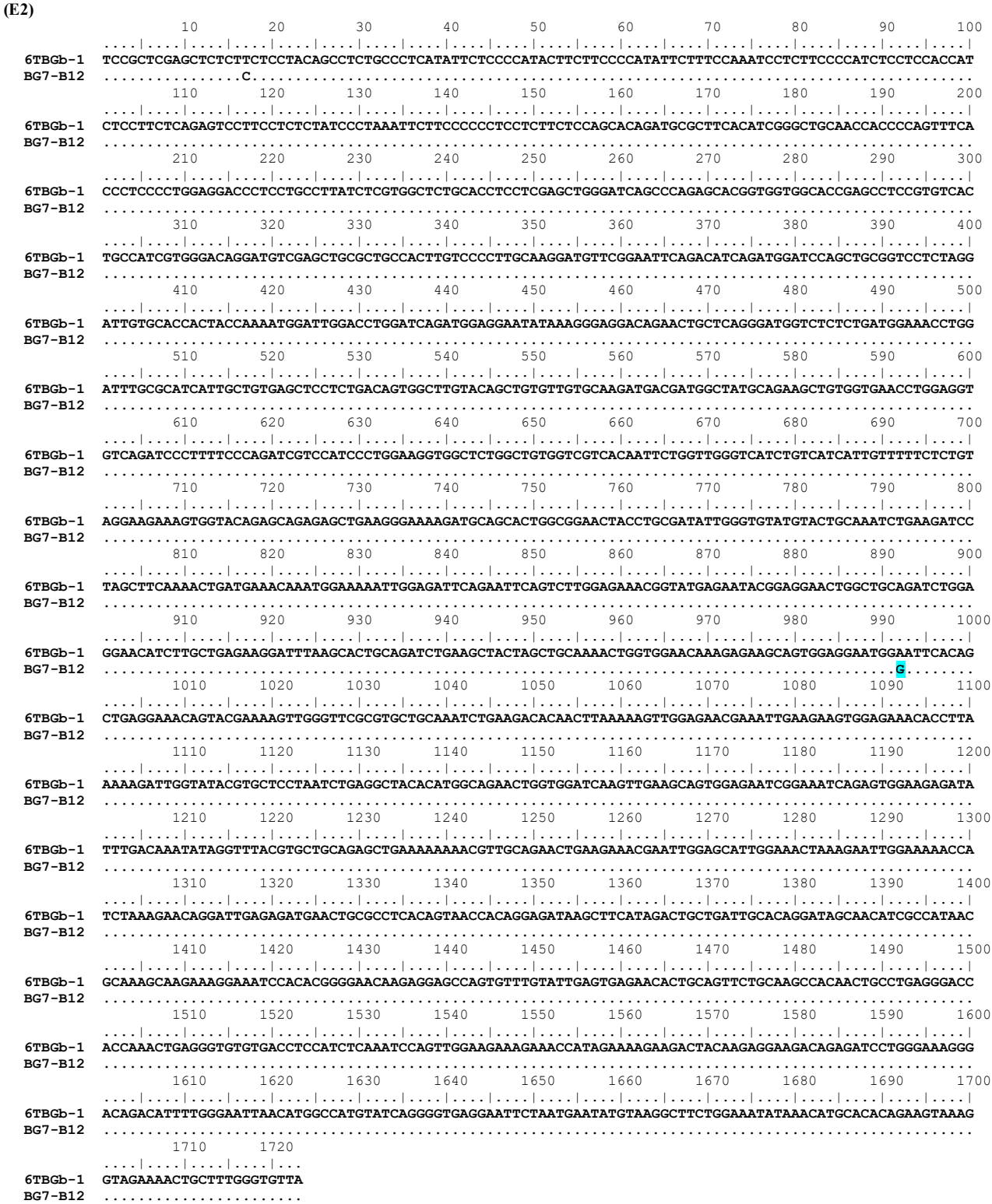
Note: PCR-based mutations in NTBGa-1 are likely to have generated NTBGa-2 (positions 695 and 1232), NTBGa-8 (position 795) and NTBGc (positions 242, 542 and insertion at position 906), and in NTBGa-4 to generate NTBGb (positions 173 and 1666).

Appendix E.

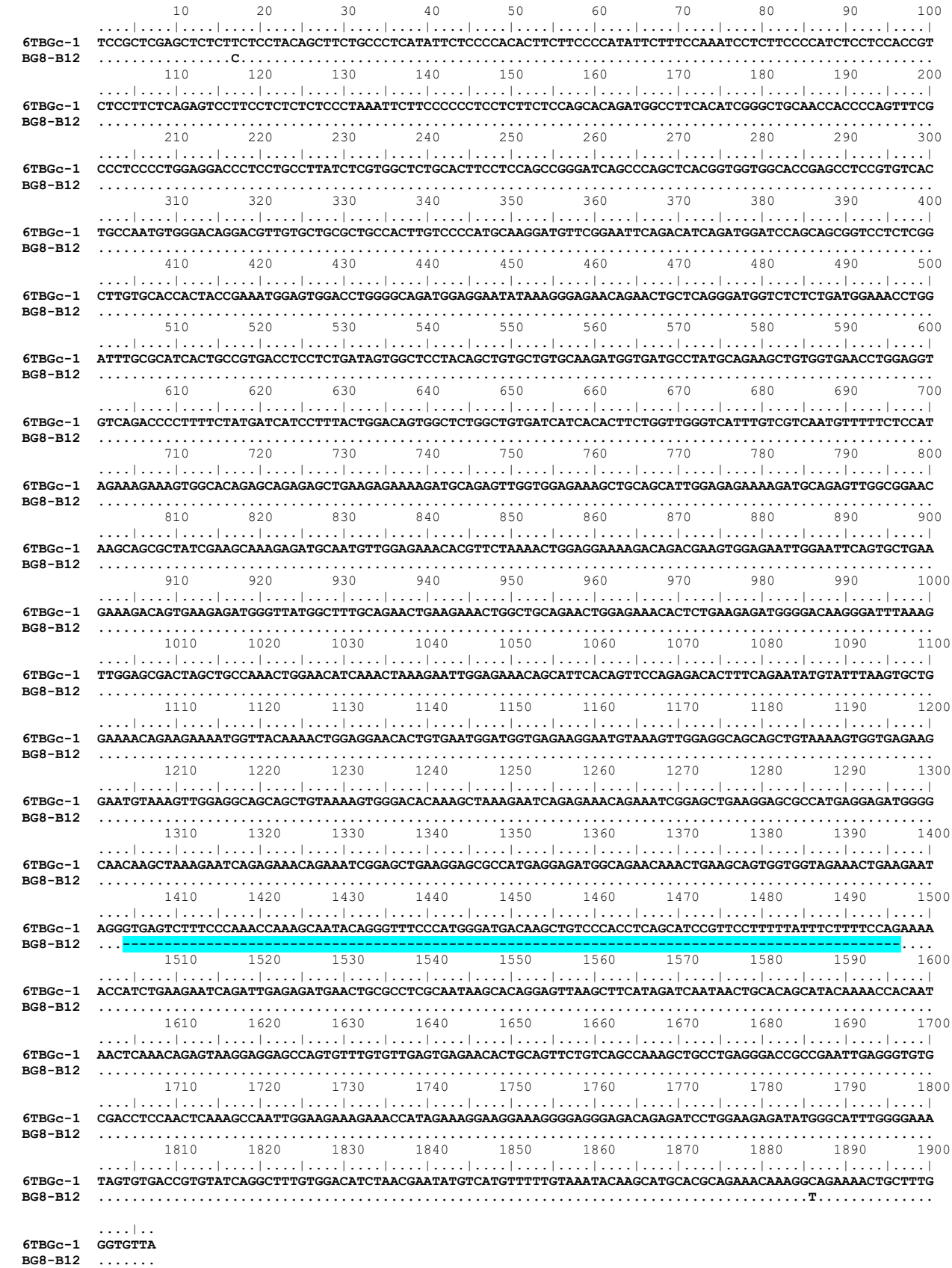
(E1)



(E2)


10 20 30 40 50 60 70 80 90 100
6TBGb-1 TCCGCTCGAGCTCTCTTCTCCTACAGCCTCTGCCCTCATATTCTCCCCATACTTCTTCCCCATATTCTTTCCAAATCCTCTTCCCCATCTCCTCCACCAT
BG7-B12 .....C.....
110 120 130 140 150 160 170 180 190 200
6TBGb-1 CTCCTTCTCAGAGTCCTTCCTCTCTATCCCTAAATTCTTCCCCCCTCCTCTTCTCCAGCACAGATGCGCTTCACATCGGGCTGCAACCACCCCAGTTTCA
BG7-B12 .....
210 220 230 240 250 260 270 280 290 300
6TBGb-1 CCCTCCCCTGGAGGACCCTCCTGCCTTATCTCGTGGCTCTGCACCTCCTCGAGCTGGGATCAGCCCAGAGCACGGTGGTGGCACCGAGCCTCCGTGTCAC
BG7-B12 .....
310 320 330 340 350 360 370 380 390 400
6TBGb-1 TGCCATCGTGGGACAGGATGTCGAGCTGCGCTGCCACTTGTCCCCTTGCAAGGATGTTCGGAATTCAGACATCAGATGGATCCAGCTGCGGTCCTCTAGG
BG7-B12 .....
410 420 430 440 450 460 470 480 490 500
6TBGb-1 ATTGTGCACCACTACCAAAATGGATTGGACCTGGATCAGATGGAGGAATATAAAGGGAGGACAGAACTGCTCAGGGATGGTCTCTCTGATGGAAACCTGG
BG7-B12 .....
510 520 530 540 550 560 570 580 590 600
6TBGb-1 ATTTGCGCATCATTGCTGTGAGCTCCTCTGACAGTGGCTTGTACAGCTGTGTTGTGCAAGATGACGATGGCTATGCAGAAGCTGTGGTGAACCTGGAGGT
BG7-B12 .....
610 620 630 640 650 660 670 680 690 700
6TBGb-1 GTCAGATCCCTTTTCCCAGATCGTCCATCCCTGGAAGGTGGCTCTGGCTGTGGTCGTCACAATTCTGGTTGGGTCATCTGTCATCATTGTTTTTCTCTGT
BG7-B12 .....
710 720 730 740 750 760 770 780 790 800
6TBGb-1 AGGAAGAAAGTGGTACAGAGCAGAGAGCTGAAGGGAAAAGATGCAGCACTGGCGGAACTACCTGCGATATTGGGTGTATGTACTGCAAATCTGAAGATCC
BG7-B12 .....
810 820 830 840 850 860 870 880 890 900
6TBGb-1 TAGCTTCAAAACTGATGAAACAAATGGAAAAATTGGAGATTCAGAATTCAGTCTTGGAGAAACGGTATGAGAATACGGAGGAACTGGCTGCAGATCTGGA
BG7-B12 .....
910 920 930 940 950 960 970 980 990 1000
6TBGb-1 GGAACATCTTGCTGAGAAGGATTTAAGCACTGCAGATCTGAAGCTACTAGCTGCAAAACTGGTGGAACAAAGAGAAGCAGTGGAGGAATGGAATTCACAG
BG7-B12 .....G.....
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
6TBGb-1 CTGAGGAAACAGTACGAAAAGTTGGGTTCGCGTGCTGCAAATCTGAAGACACAACTTAAAAAGTTGGAGAACGAAATTGAAGAAGTGGAGAAACACCTTA
BG7-B12 .....
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
6TBGb-1 AAAAGATTGGTATACGTGCTCCTAATCTGAGGCTACACATGGCAGAACTGGTGGATCAAGTTGAAGCAGTGGAGAATCGGAAATCAGAGTGGAAGAGATA
BG7-B12 .....
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
6TBGb-1 TTTGACAAATATAGGTTTACGTGCTGCAGAGCTGAAAAAAAACGTTGCAGAACTGAAGAAACGAATTGGAGCATTGGAAACTAAAGAATTGGAAAAACCA
BG7-B12 .....
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
6TBGb-1 TCTAAAGAACAGGATTGAGAGATGAACTGCGCCTCACAGTAACCACAGGAGATAAGCTTCATAGACTGCTGATTGCACAGGATAGCAACATCGCCATAAC
BG7-B12 .....
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
6TBGb-1 GCAAAGCAAGAAAGGAAATCCACACGGGGAACAAGAGGAGCCAGTGTTTGTATTGAGTGAGAACACTGCAGTTCTGCAAGCCACAACTGCCTGAGGGACC
BG7-B12 .....
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
6TBGb-1 ACCAAACTGAGGGTGTGTGACCTCCATCTCAAATCCAGTTGGAAGAAAGAAACCATAGAAAAGAAGACTACAAGAGGAAGACAGAGATCCTGGGAAAGGG
BG7-B12 .....
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
6TBGb-1 ACAGACATTTTGGGAATTAACATGGCCATGTATCAGGGGTGAGGAATTCTAATGAATATGTAAGGCTTCTGGAAATATAAACATGCACACAGAAGTAAAG
BG7-B12 .....
1710 1720
6TBGb-1 GTAGAAAACTGCTTTGGGTGTTA
BG7-B12 .....

(E3)

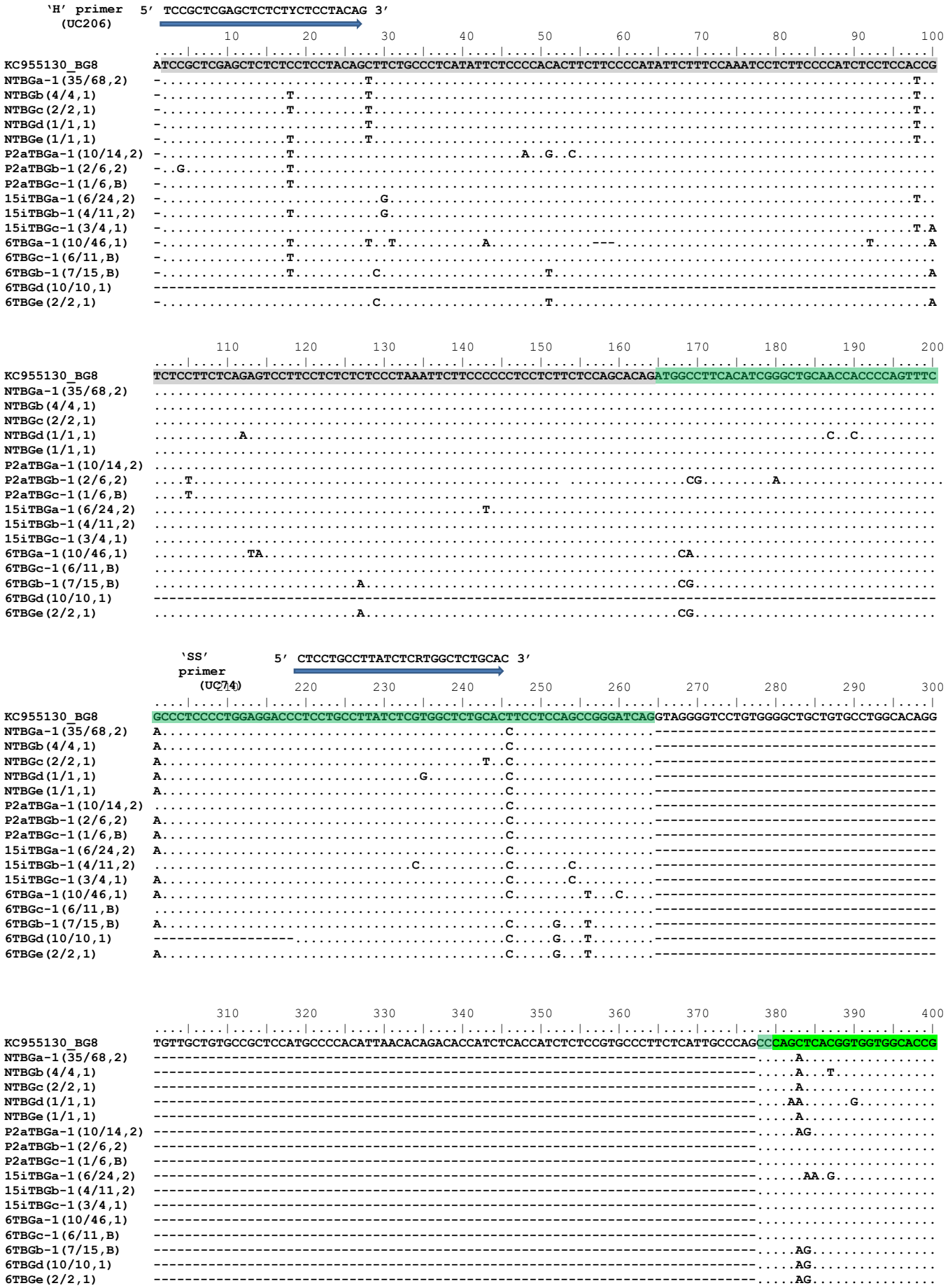

10 20 30 40 50 60 70 80 90 100
6TBGc-1 TCCGCTCGAGCTCTCTTCTCCTACAGCTTCTGCCCTCATATTCTCCCCACACTTCTTCCCCATATTCTTTCCAAATCCTCTTCCCCATCTCCTCCACCGT
BG8-B12 .....C.....
110 120 130 140 150 160 170 180 190 200
6TBGc-1 CTCCTTCTCAGAGTCCTTCCTCTCTCTCCCTAAATTCTTCCCCCCTCCTCTTCTCCAGCACAGATGGCCTTCACATCGGGCTGCAACCACCCCAGTTTCG
BG8-B12 .....
210 220 230 240 250 260 270 280 290 300
6TBGc-1 CCCTCCCCTGGAGGACCCTCCTGCCTTATCTCGTGGCTCTGCACTTCCTCCAGCCGGGATCAGCCCAGCTCACGGTGGTGGCACCGAGCCTCCGTGTCAC
BG8-B12 .....
310 320 330 340 350 360 370 380 390 400
6TBGc-1 TGCCAATGTGGGACAGGACGTTGTGCTGCGCTGCCACTTGTCCCCATGCAAGGATGTTCGGAATTCAGACATCAGATGGATCCAGCAGCGGTCCTCTCGG
BG8-B12 .....
410 420 430 440 450 460 470 480 490 500
6TBGc-1 CTTGTGCACCACTACCGAAATGGAGTGGACCTGGGGCAGATGGAGGAATATAAAGGGAGAACAGAACTGCTCAGGGATGGTCTCTCTGATGGAAACCTGG
BG8-B12 .....
510 520 530 540 550 560 570 580 590 600
6TBGc-1 ATTTGCGCATCACTGCCGTGACCTCCTCTGATAGTGGCTCCTACAGCTGTGCTGTGCAAGATGGTGATGCCTATGCAGAAGCTGTGGTGAACCTGGAGGT
BG8-B12 .....
610 620 630 640 650 660 670 680 690 700
6TBGc-1 GTCAGACCCCTTTTCTATGATCATCCTTTACTGGACAGTGGCTCTGGCTGTGATCATCACACTTCTGGTTGGGTCATTTGTCGTCAATGTTTTTCTCCAT
BG8-B12 .....
710 720 730 740 750 760 770 780 790 800
6TBGc-1 AGAAAGAAAGTGGCACAGAGCAGAGAGCTGAAGAGAAAAGATGCAGAGTTGGTGGAGAAAGCTGCAGCATTGGAGAGAAAAGATGCAGAGTTGGCGGAAC
BG8-B12 .....
810 820 830 840 850 860 870 880 890 900
6TBGc-1 AAGCAGCGCTATCGAAGCAAAGAGATGCAATGTTGGAGAAACACGTTCTAAAACTGGAGGAAAAGACAGACGAAGTGGAGAATTGGAATTCAGTGCTGAA
BG8-B12 .....
910 920 930 940 950 960 970 980 990 1000
6TBGc-1 GAAAGACAGTGAAGAGATGGGTTATGGCTTTGCAGAACTGAAGAAACTGGCTGCAGAACTGGAGAAACACTCTGAAGAGATGGGGACAAGGGATTTAAAG
BG8-B12 .....
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
6TBGc-1 TTGGAGCGACTAGCTGCCAAACTGGAACATCAAACTAAAGAATTGGAGAAACAGCATTCACAGTTCCAGAGACACTTTCAGAATATGTATTTAAGTGCTG
BG8-B12 .....
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
6TBGc-1 GAAAACAGAAGAAAATGGTTACAAAACTGGAGGAACACTGTGAATGGATGGTGAGAAGGAATGTAAAGTTGGAGGCAGCAGCTGTAAAAGTGGTGAGAAG
BG8-B12 .....
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
6TBGc-1 GAATGTAAAGTTGGAGGCAGCAGCTGTAAAAGTGGGACACAAAGCTAAAGAATCAGAGAAACAGAAATCGGAGCTGAAGGAGCGCCATGAGGAGATGGGG
BG8-B12 .....
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
6TBGc-1 CAACAAGCTAAAGAATCAGAGAAACAGAAATCGGAGCTGAAGGAGCGCCATGAGGAGATGGCAGAACAAACTGAAGCAGTGGTGGTAGAAACTGAAGAAT
BG8-B12 .....
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
6TBGc-1 AGGGTGAGTCTTTCCCAAACCAAAGCAATACAGGGTTTCCCATGGGATGACAAGCTGTCCCACCTCAGCATCCGTTCCTTTTTATTTCTTTTCCAGAAAA
BG8-B12 ...----------.....
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
6TBGc-1 ACCATCTGAAGAATCAGATTGAGAGATGAACTGCGCCTCGCAATAAGCACAGGAGTTAAGCTTCATAGATCAATAACTGCACAGCATACAAAACCACAAT
BG8-B12 .....
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
6TBGc-1 AACTCAAACAGAGTAAGGAGGAGCCAGTGTTTGTGTTGAGTGAGAACACTGCAGTTCTGTCAGCCAAAGCTGCCTGAGGGACCGCCGAATTGAGGGTGTG
BG8-B12 .....
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
6TBGc-1 CGACCTCCAACTCAAAGCCAATTGGAAGAAAGAAACCATAGAAAGGAAGGAAAGGGGAGGGAGACAGAGATCCTGGAAGAGATATGGGCATTTGGGGAAA
BG8-B12 .....
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
6TBGc-1 TAGTGTGACCGTGTATCAGGCTTTGTGGACATCTAACGAATATGTCATGTTTTTGTAAATACAAGCATGCACGCAGAAACAAAGGCAGAAAACTGCTTTG
BG8-B12 .....T.....
....|..
6TBGc-1 GGTGTTA
BG8-B12 .....

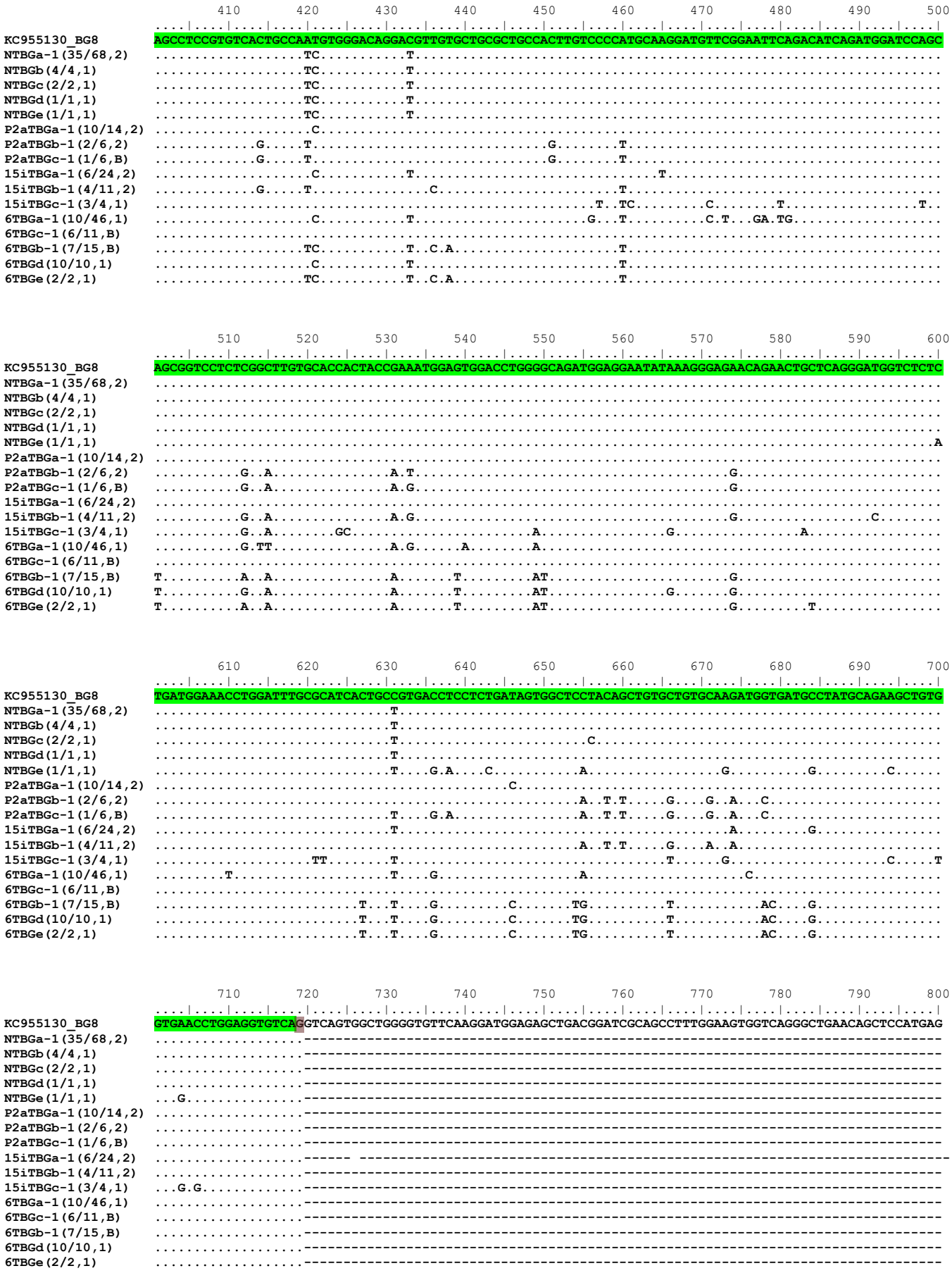
(E4)

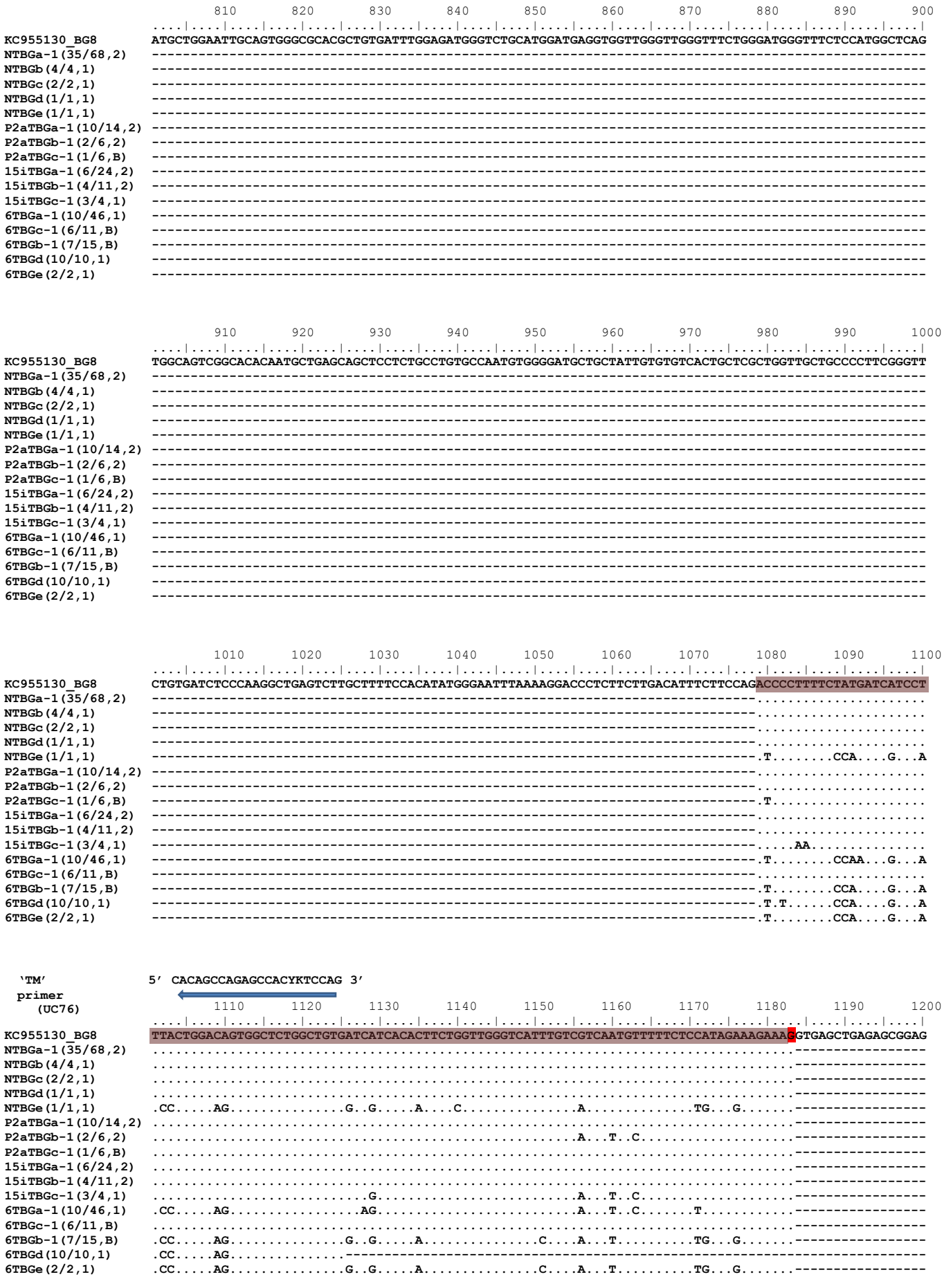
```
                10        20        30        40        50        60        70        80        90       100
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6TBGd     ----------------------------------------------------------------------------------------------------
BG11-B12 ATCCGCTCGAGCTCTCTCCTCCTACAGCCTCTGCCCTCATATTCTCCCCATACTTCTTCCCCATATTCTTTCCAAATCCTCTTCCCCATCTCCTCCACCA
                110       120       130       140       150       160       170       180       190       200
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6TBGd     ----------------------------------------------------------------------------------------------------
BG11-B12 TCTCCTTCTCAGAGTCCTTCCTCTCTCTCCCTAAATTCTTCCCCCCTCCTCTTCTCCAGCACAGATGCGCTTCACATCGGGATGCAACCACCCCAGATTC
                210       220       230       240       250       260       270       280       290       300
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6TBGd     ------------------CTCCTGCCTTATCTCGTGGCTCTGCACCTCCTCGAGCTGGGATCAGCCCAGAGCACGGTGGTGGCACCGAGCCTCCGTGTCA
BG11-B12 ACCCTCCCCTGGAGGACC..................................................................................
                310       320       330       340       350       360       370       380       390       400
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6TBGd     CTGCCAACGTGGGACAGGATGTTGTGCTGCGCTGCCACTTGTCCCCTTGCAAGGATGTTCGGAATTCAGACATCAGATGGATCCAGCTGCGGTCCTCTGG
BG11-B12 ..................................................................................................
                410       420       430       440       450       460       470       480       490       500
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6TBGd     GATTGTGCACCACTACCAAAATGGATTGGACCTGGATCAGATGGAGGAATATGAAGGGAGGACAGAACTGCTCAGGGATGGTCTCTCTGATGGAAACCTG
BG11-B12 ..................................................................................................
                510       520       530       540       550       560       570       580       590       600
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6TBGd     GATTTGCGCATCATTGCTGTGAGCTCCTCTGACAGTGGCTTGTACAGCTGTGTTGTGCAAGATGACGATGGCTATGCAGAAGCTGTGGTGAACCTGGAGG
BG11-B12 ..................................................................................................
                610       620       630       640       650       660       670       680       690       700
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6TBGd     TGTCAGATCTCTTTTCCCAGATCGTCCATCCCTGGAAGGTGGCTCTGGCTGT-----------------------------------------------
BG11-B12 ..................................................GGTCGTCACAATTCTGGTTGGGTCATTTGTCATCATTGCTTTTCTCTA
                710
          ....|....|
6TBGd     ----------
BG11-B12 TAGGAAGAAA
```

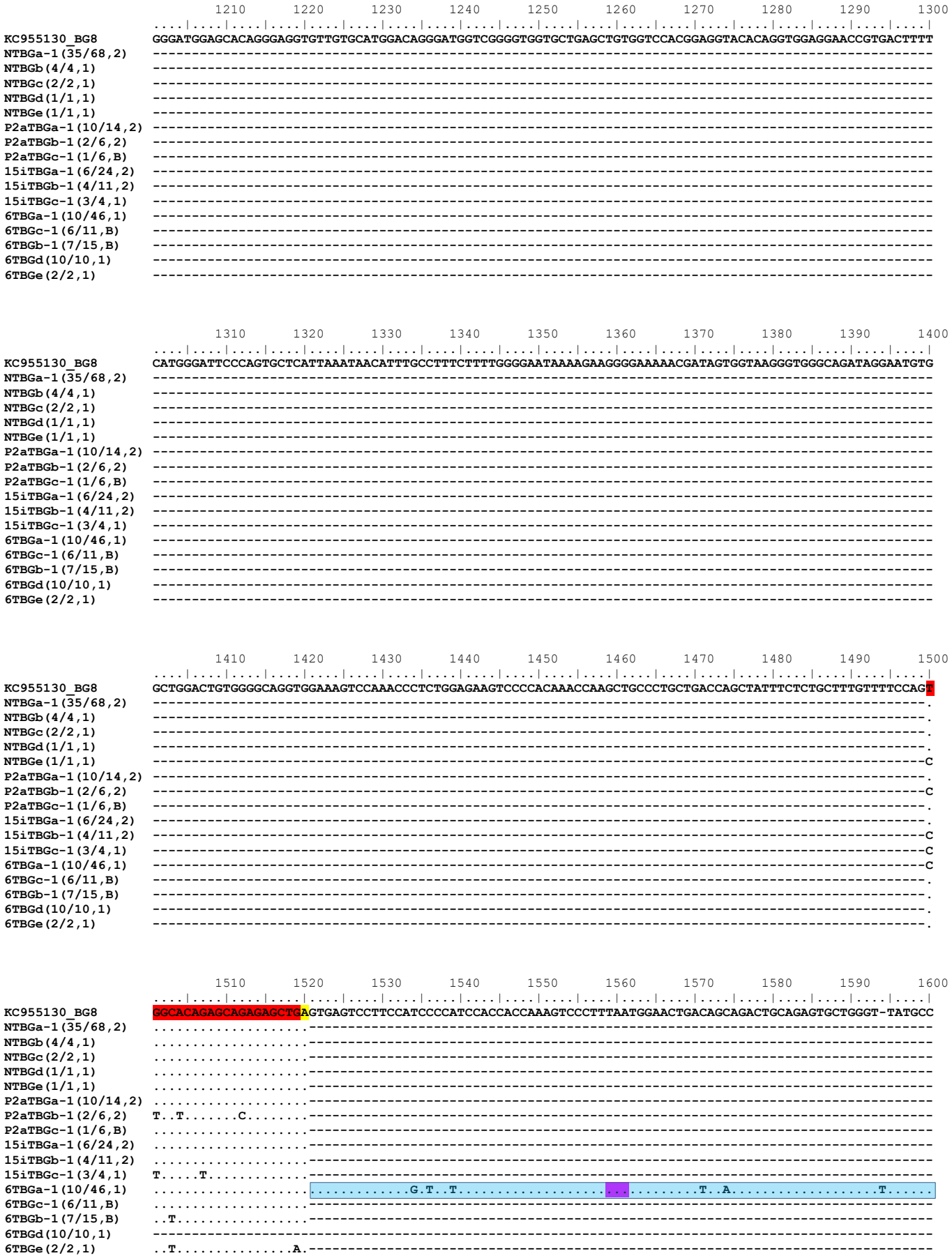
Appendix E. Nucleotide sequence alignment shows that four BG genes from T cells of line 6_1 (B2) are the same genes as the ones in B12 haplotype. Names of the transcripts follow the convention: abbreviated line name, "T" for T cells, "BG", a letter representing the exon 2 sequence with "a" being the most frequently detected exon 2 sequence (and "b" being the second most frequently detected exon 2 sequence, and so forth), a dash and then a number representing the alternative splicing variant with "1" being the most frequently detected clone. Letters indicate nucleotides, dots indicate identities with the top sequences, dashes indicate no sequence present compared to the top sequences. The differences between the major cDNA sequences from the four BG genes of B2 and their counterparts from B12 are labeled with blue colour. The differences were either due to one nucleotide PCR artifact in (E2) or the intron retention in (E1) and (E3). It should be noted that the B12 BG cDNA sequences were annotated from genomic sequences, which might or might not represent the expressed transcripts. Note: 6TBGd only contains partial cDNA sequence (partial signal sequence, the whole Ig-V domain and partial transmembrane region) as it was found using SS-TM primers.

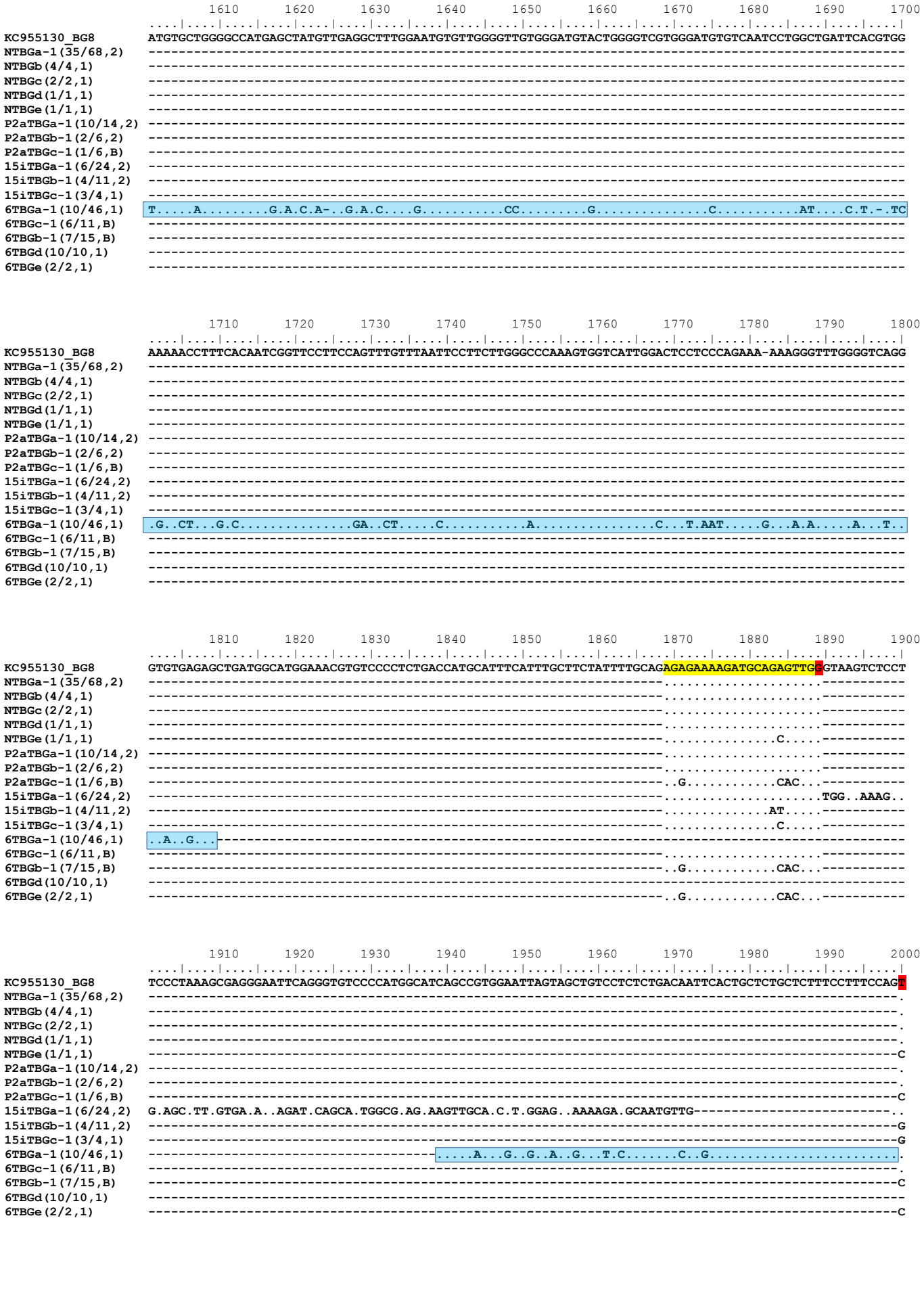
Appendix F.

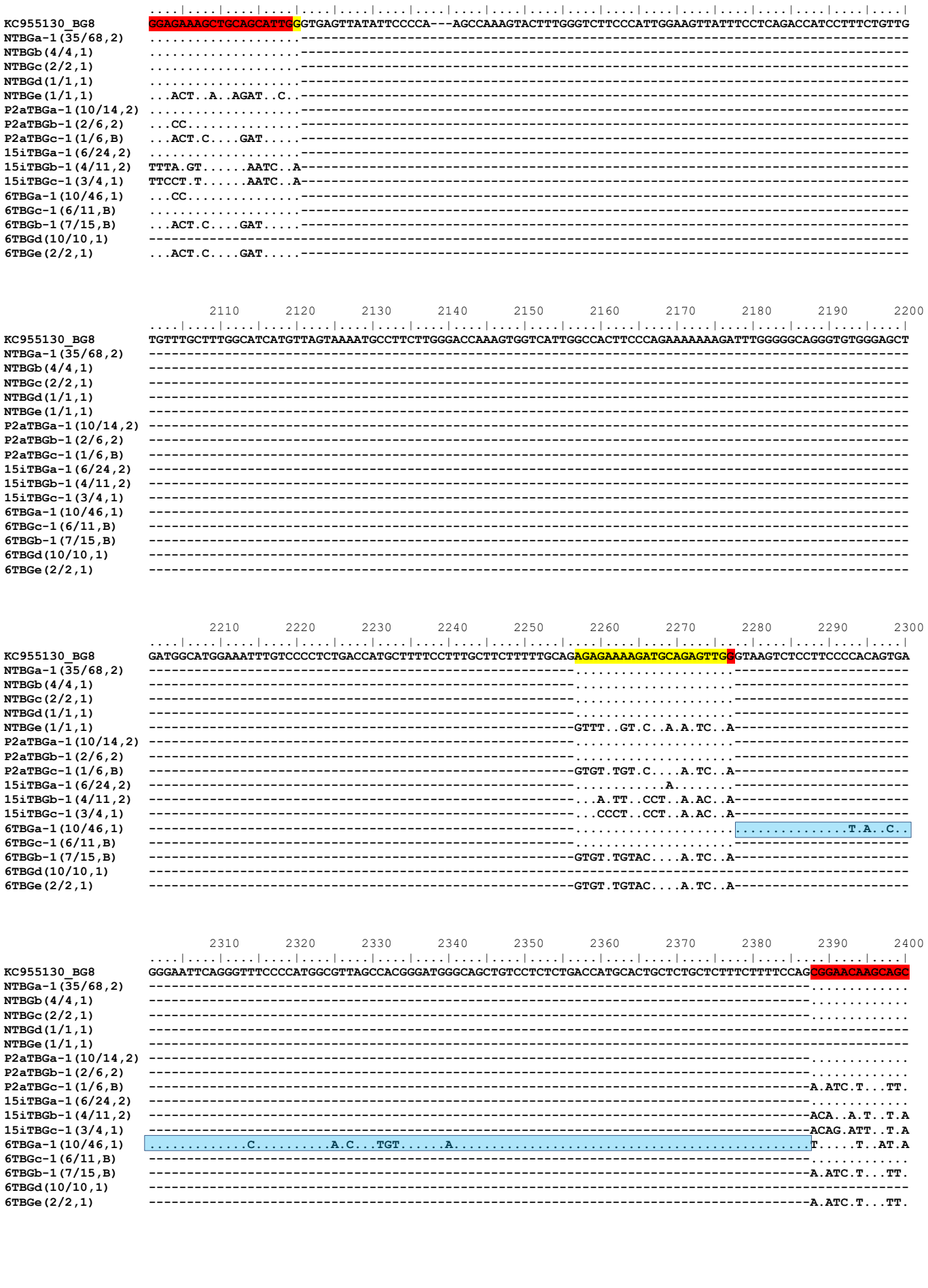

'H' primer (UC206) 5' TCCGCTCGAGCTCTCTYCTCCTACAG 3'
'SS' primer (UC214) 5' CTCCTGCCTTATCTCRTGGCTCTGCAC 3'
KC955130_BG8
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)

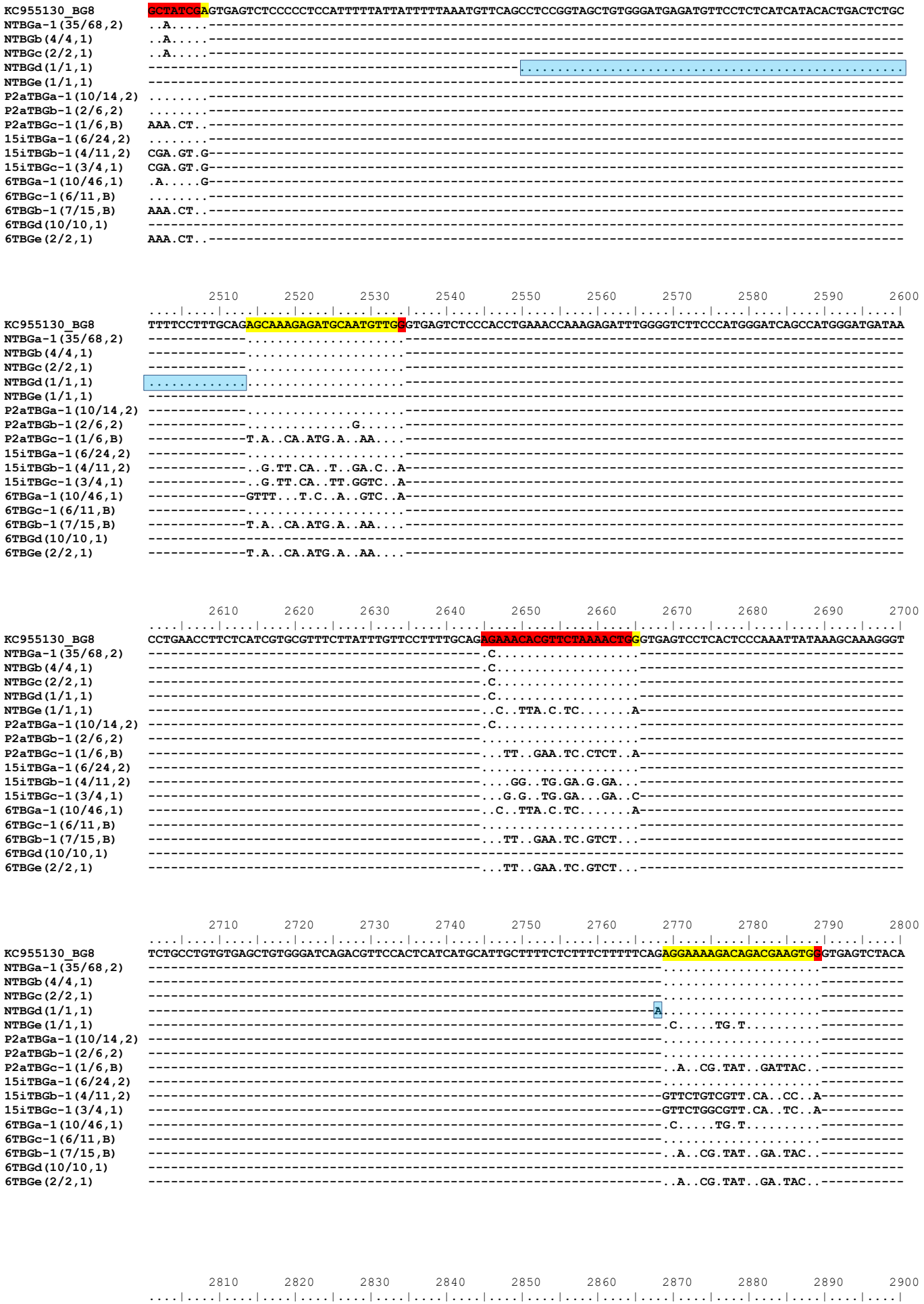

410 420 430 440 450 460 470 480 490 500
KC955130_BG8 AGCCTCCGTGTCACTGCCAATGTGGGACAGGACGTTGTGCTGCGCTGCCACTTGTCCCCATGCAAGGATGTTCGGAATTCAGACATCAGATGGATCCAGC
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
510 520 530 540 550 560 570 580 590 600
KC955130_BG8 AGCGGTCCTCTCGGCTTGTGCACCACTACCGAAATGGAGTGGACCTGGGGCAGATGGAGGAATATAAAGGGAGAACAGAACTGCTCAGGGATGGTCTCTC
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
610 620 630 640 650 660 670 680 690 700
KC955130_BG8 TGATGGAAACCTGGATTTGCGCATCACTGCCGTGACCTCCTCTGATAGTGGCTCCTACAGCTGTGCTGTGCAAGATGGTGATGCCTATGCAGAAGCTGTG
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
710 720 730 740 750 760 770 780 790 800
KC955130_BG8 GTGAACCTGGAGGTGTCAGGTCAGTGGCTGGGGTGTTCAAGGATGGAGAGCTGACGGATCGCAGCCTTTGGAAGTGGTCAGGGCTGAACAGCTCCATGAG
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)

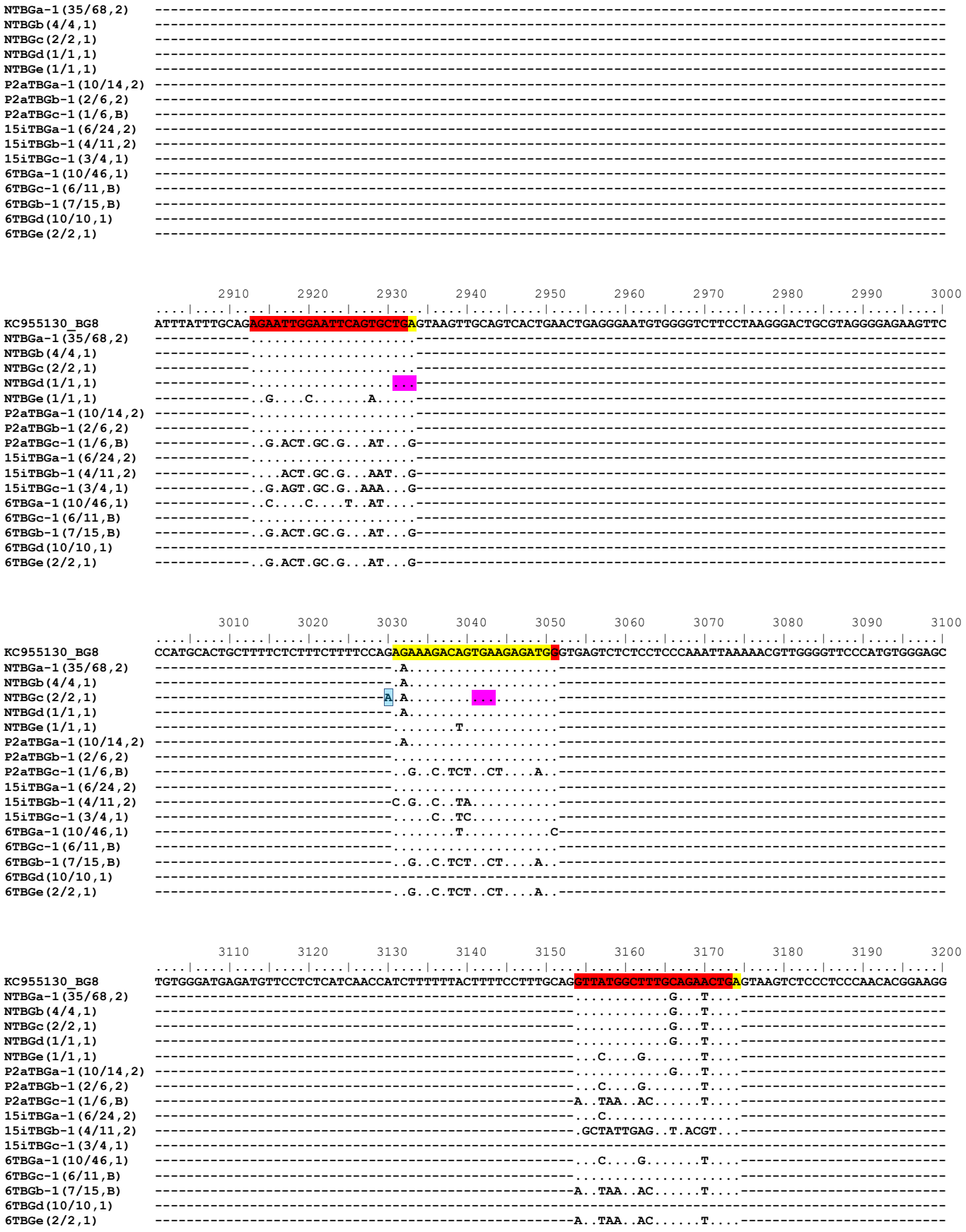

810 820 830 840 850 860 870 880 890 900
KC955130_BG8 ATGCTGGAATTGCAGTGGGCGCACGCTGTGATTTGGAGATGGGTCTGCATGGATGAGGTGGTTGGGTTGGGTTTCTGGGATGGGTTTCTCCATGGCTCAG
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
910 920 930 940 950 960 970 980 990 1000
KC955130_BG8 TGGCAGTCGGCACACAATGCTGAGCAGCTCCTCTGCCTGTGCCAATGTGGGGATGCTGCTATTGTGTGTCACTGCTCGCTGGTTGCTGCCCCTTCGGGTT
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
KC955130_BG8 CTGTGATCTCCCAAGGCTGAGTCTTGCTTTTCCACATATGGGAATTTAAAAGGACCCTCTTCTTGACATTTCTTCCAGACCCCTTTTCTATGATCATCCT
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1) .T.....CCA....G...A
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B) .T.....
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1) .....AA.....
6TBGa-1(10/46,1) .T.....CCAA...G...A
6TBGc-1(6/11,B)
6TBGb-1(7/15,B) .T.....CCA....G...A
6TBGd(10/10,1) .T.T.....CCA....G...A
6TBGe(2/2,1) .T.....CCA....G...A
'TM' primer (UC76) 5′ CACAGCCAGAGCCACYKTCCAG 3′
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
KC955130_BG8 TTACTGGACAGTGGCTCTGGCTGTGATCATCACACTTCTGGTTGGGTCATTTGTCGTCAATGTTTTTCTCCATAGAAAGAAAAGGTGAGCTGAGAGCGGAG
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1) .CC.....AG.....G..G.....A....C.....A.....TG...G.....
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2) .....A...T..C.....
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1) .....G.....A...T..C.....
6TBGa-1(10/46,1) .CC.....AG.....AG.....A...T..C.....T.....
6TBGc-1(6/11,B)
6TBGb-1(7/15,B) .CC.....AG.....G..G.....A.....C....A...T.....TG...G.....
6TBGd(10/10,1) .CC.....AG.....
6TBGe(2/2,1) .CC.....AG.....G..G.....A.....C....A...T.....TG...G.....

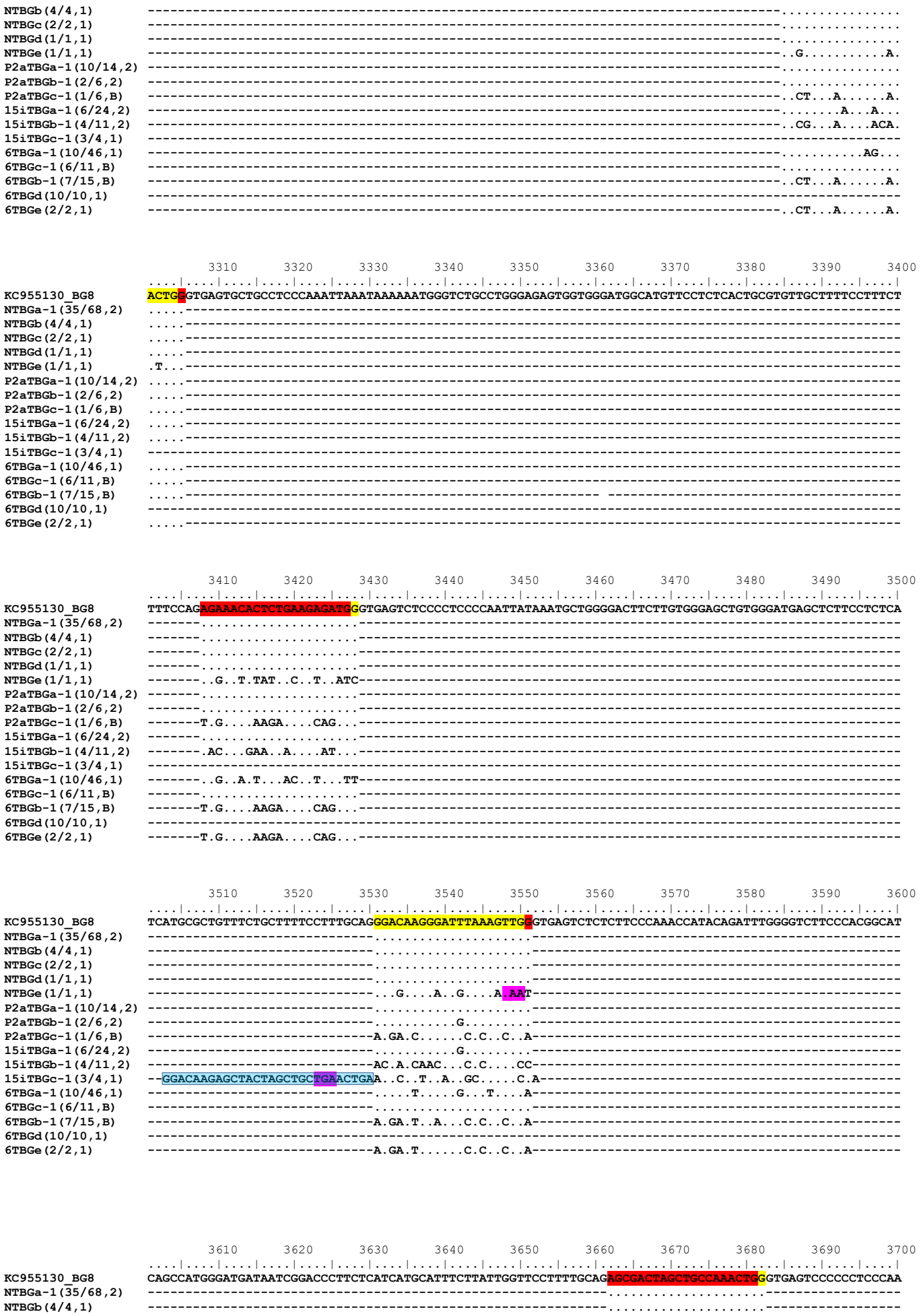

1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
KC955130_BG8 GGGATGGAGCACAGGGAGGTGTTGTGCATGGACAGGGATGGTCGGGGTGGTGCTGAGCTGTGGTCCACGGAGGTACACAGGTGGAGGAACCGTGACTTTT
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
KC955130_BG8 CATGGGATTCCCAGTGCTCATTAAATAACATTTGCCTTTCTTTTGGGGAATAAAAGAAGGGGAAAAACGATAGTGGTAAGGGTGGGCAGATAGGAATGTG
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
KC955130_BG8 GCTGGACTGTGGGGCAGGTGGAAAGTCCAAACCCTCTGGAGAAGTCCCCACAAACCAAGCTGCCCTGCTGACCAGCTATTTCTCTGCTTTGTTTTCCAGT
NTBGa-1(35/68,2) .
NTBGb(4/4,1) .
NTBGc(2/2,1) .
NTBGd(1/1,1) .
NTBGe(1/1,1) C
P2aTBGa-1(10/14,2) .
P2aTBGb-1(2/6,2) C
P2aTBGc-1(1/6,B) .
15iTBGa-1(6/24,2) .
15iTBGb-1(4/11,2) C
15iTBGc-1(3/4,1) C
6TBGa-1(10/46,1) C
6TBGc-1(6/11,B) .
6TBGb-1(7/15,B) .
6TBGd(10/10,1) .
6TBGe(2/2,1) .
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
KC955130_BG8 GGCACAGAGCAGAGAGCTGAGTGAGTCCTTCCATCCCCATCCACCACCAAAGTCCCTTTAATGGAACTGACAGCAGACTGCAGAGTGCTGGGT-TATGCC
NTBGa-1(35/68,2) .....
NTBGb(4/4,1) .....
NTBGc(2/2,1) .....
NTBGd(1/1,1) .....
NTBGe(1/1,1) .....
P2aTBGa-1(10/14,2) .....
P2aTBGb-1(2/6,2) T..T.....C.....
P2aTBGc-1(1/6,B) .....
15iTBGa-1(6/24,2) .....
15iTBGb-1(4/11,2) .....
15iTBGc-1(3/4,1) T.....T.....
6TBGa-1(10/46,1) .....G.T..T.....T..A.....T.....
6TBGc-1(6/11,B) .....
6TBGb-1(7/15,B) ..T.....
6TBGd(10/10,1)
6TBGe(2/2,1) ..T.....A.

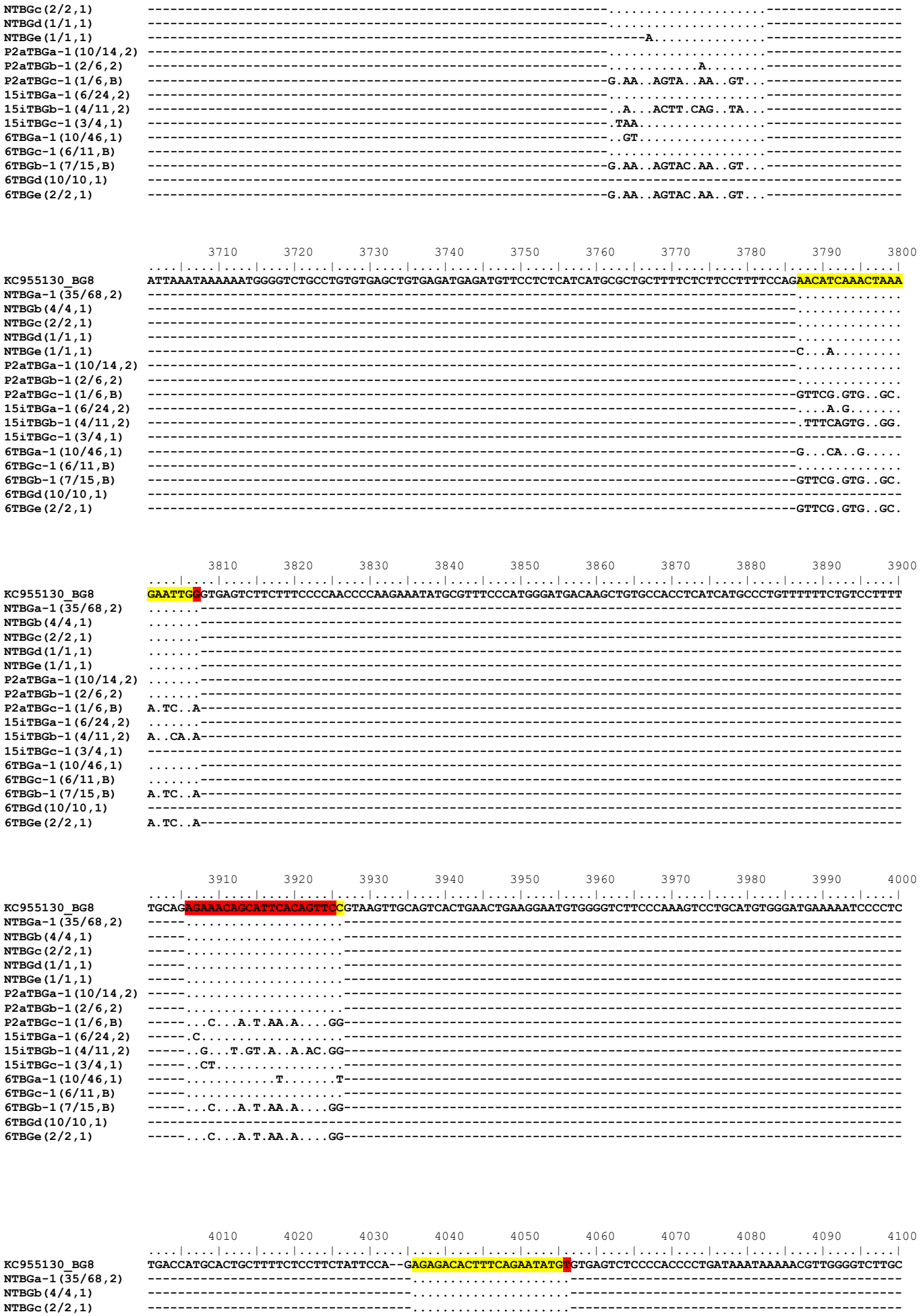

1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
KC955130_BG8 ATGTGCTGGGGCCATGAGCTATGTTGAGGCTTTGGAATGTGTTGGGGTTGTGGGATGTACTGGGGTCGTGGGATGTGTCAATCCTGGCTGATTCACGTGG
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1) T.....A.........G.A.C.A-..G.A.C....G...........CC.........G..............C...........AT....C.T.-.TC
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
KC955130_BG8 AAAAACCTTTCACAATCGGTTCCTTCCAGTTTGTTTAATTCCTTCTTGGGCCCAAAGTGGTCATTGGACTCCTCCCAGAAA-AAAGGGTTTGGGGTCAGG
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1) .G..CT...G.C..............GA..CT.....C..........A..............C...T.AAT.....G...A.A.....A...T..
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
KC955130_BG8 GTGTGAGAGCTGATGGCATGGAAACGTGTCCCCTCTGACCATGCATTTCATTTGCTTCTATTTTGCAGAGAGAAAAGATGCAGAGTTGGGTAAGTCTCCT
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1) ..A..G...
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
KC955130_BG8 TCCCTAAAGCGAGGGAATTCAGGGTGTCCCCATGGCATCAGCCGTGGAATTAGTAGCTGTCCTCTCTGACAATTCACTGCTCTGCTCTTTCCTTTCCAGT
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2) G.AGC.TT.GTGA.A..AGAT.CAGCA.TGGCG.AG.AAGTTGCA.C.T.GGAG..AAAAGA.GCAATGTTG----------..
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1) .....A...G..G..A..G...T.C.......C..G.....
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
2010 2020 2030 2040 2050 2060 2070 2080 2090 2100

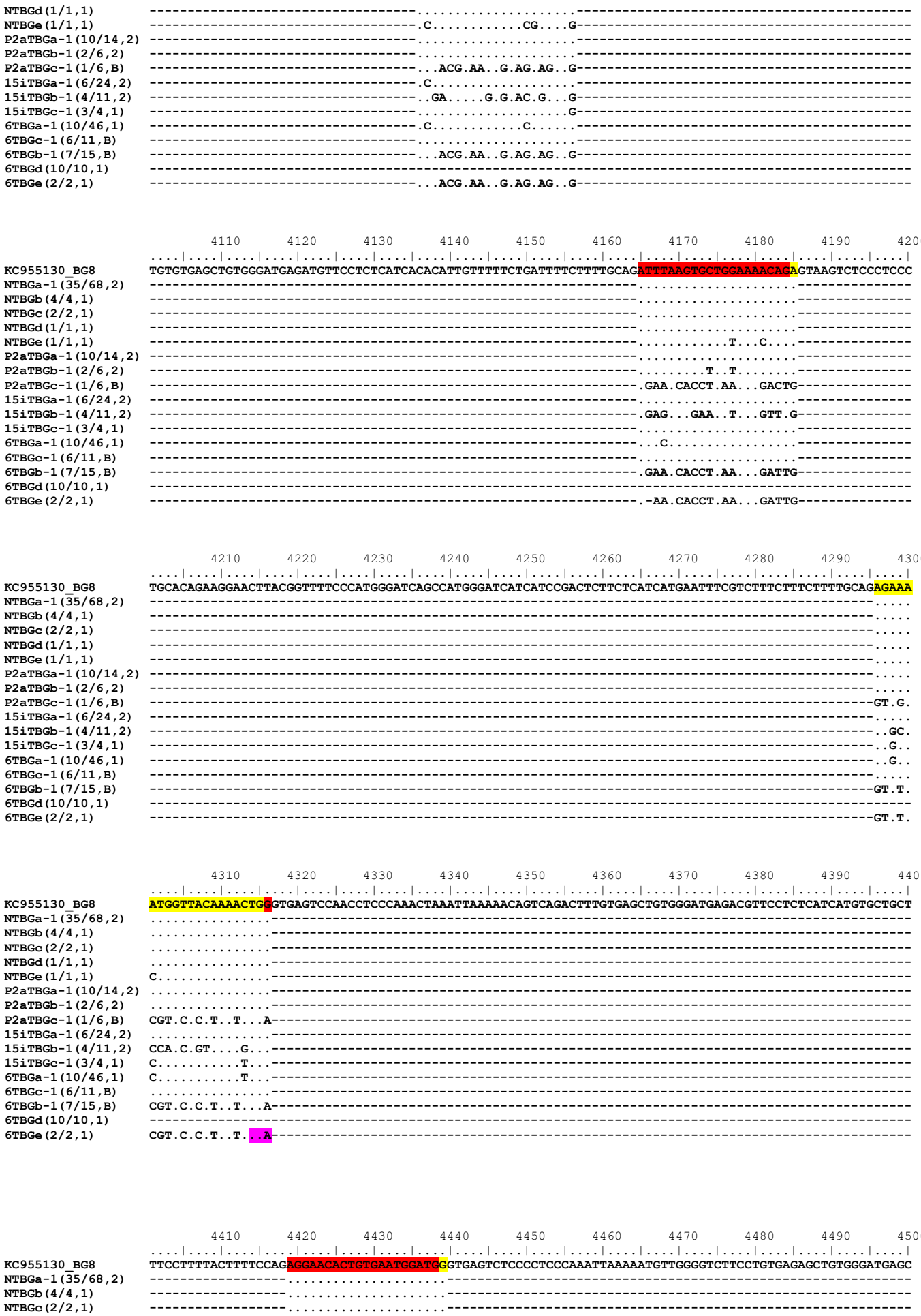

KC955130_BG8
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
2410 2420 2430 2440 2450 2460 2470 2480 2490 2500

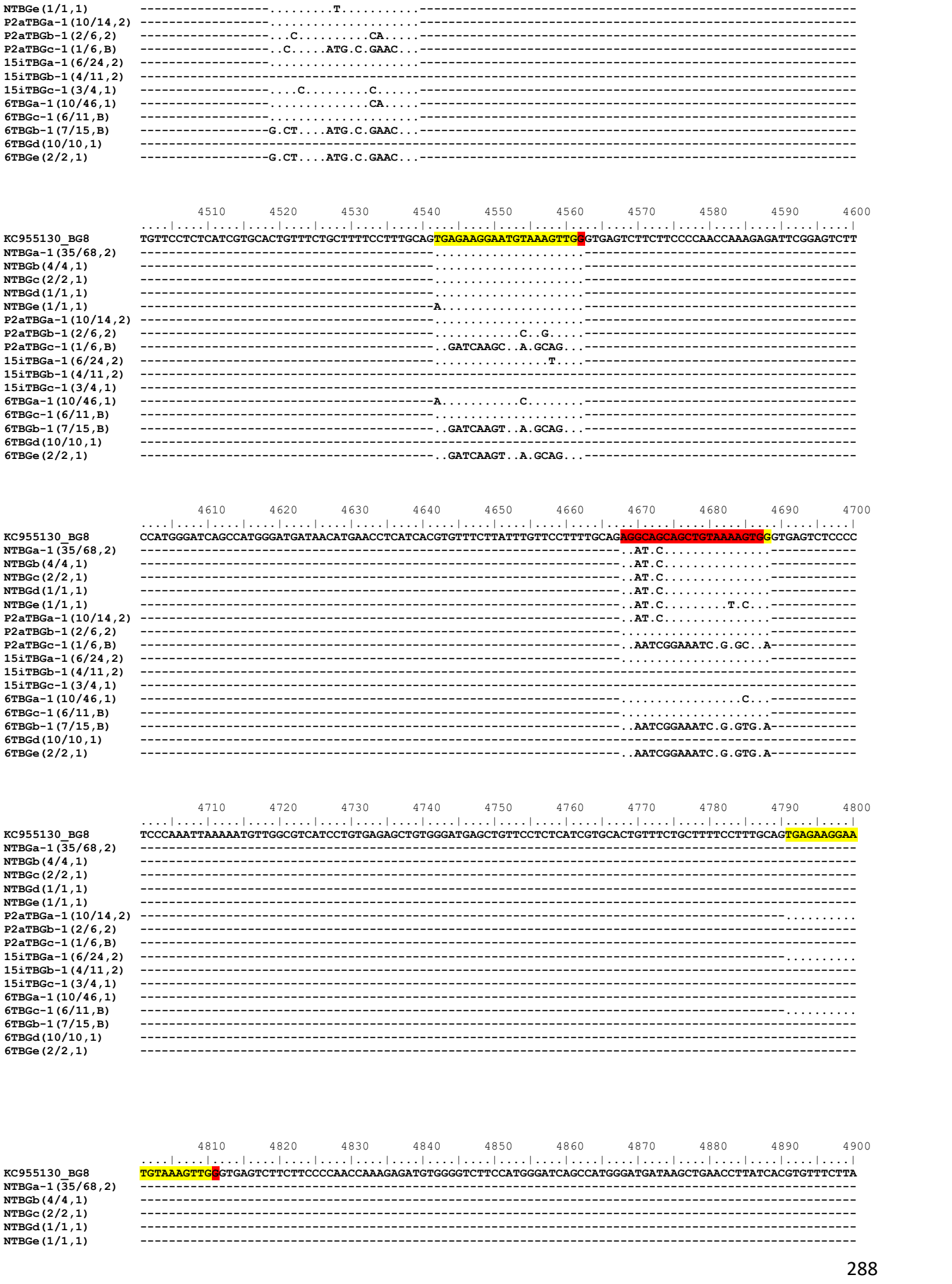

KC955130_BG8
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
2810 2820 2830 2840 2850 2860 2870 2880 2890 2900

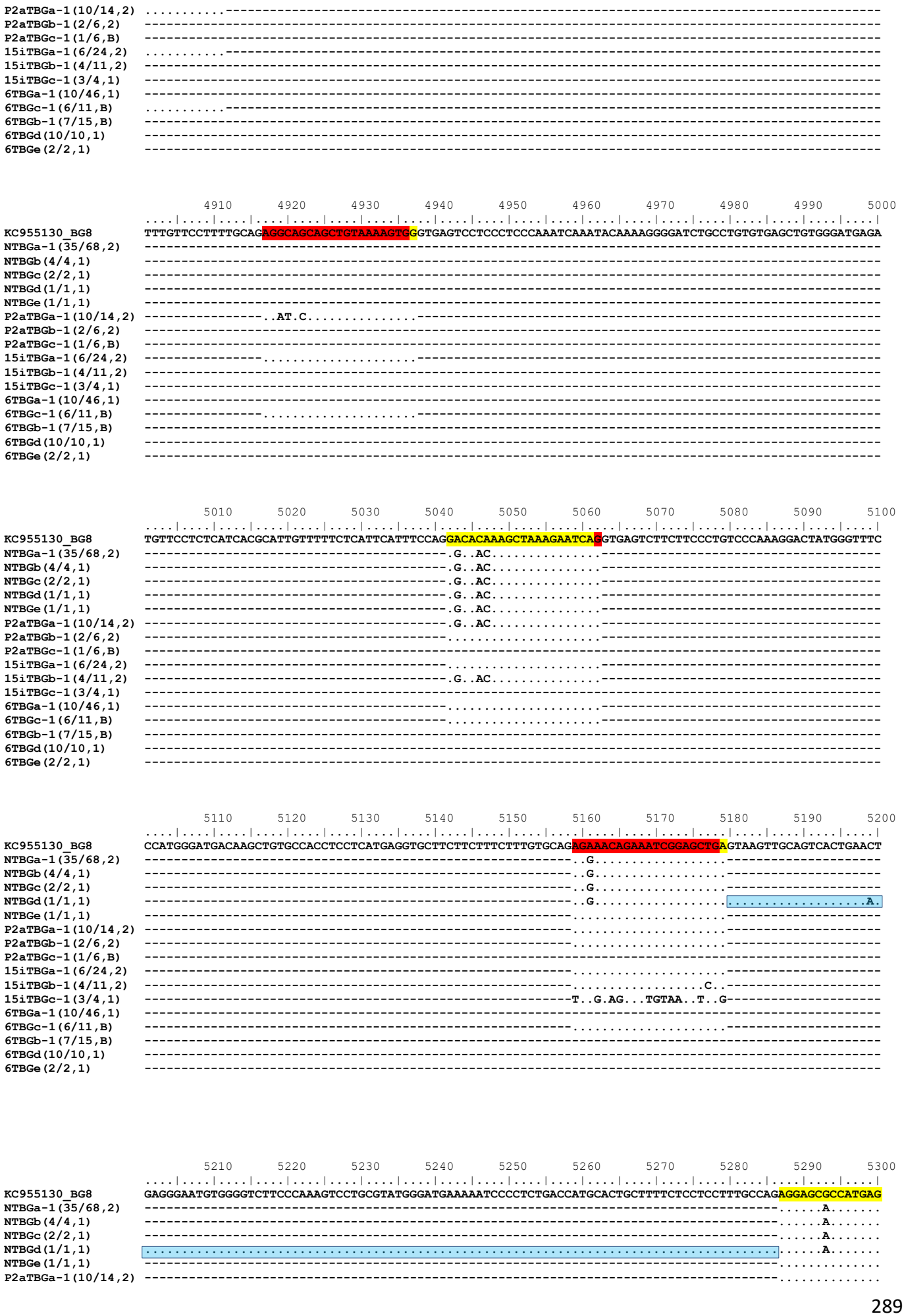

NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
KC955130_BG8 ATTTATTTGCAGAGAATTGGAATTCAGTGCTGAGTAAGTTGCAGTCACTGAACTGAGGGAATGTGGGGTCTTCCTAAGGGACTGCGTAGGGGAGAAGTTC
3010 3020 3030 3040 3050 3060 3070 3080 3090 3100
KC955130_BG8 CCATGCACTGCTTTTCTCTTTCTTTTCCAGAGAAAGACAGTGAAGAGATGGGTGAGTCTCTCCTCCCAAATTAAAAACGTTGGGGTTCCCATGTGGGAGC
3110 3120 3130 3140 3150 3160 3170 3180 3190 3200
KC955130_BG8 TGTGGGATGAGATGTTCCTCTCATCAACCATCTTTTTTACTTTTCCTTTGCAGGTTATGGCTTTGCAGAACTGAGTAAGTCTCCCTCCCAACACGGAAGG
3210 3220 3230 3240 3250 3260 3270 3280 3290 3300
KC955130_BG8 GATTTGTGGTCTTCCCATGGGATCAGCCATGGGATGATCATCTGACCCCTCTCATCATGCATTTCGTATTTGTTTCTTTTGCAGAGAAACTGGCTGCAGA

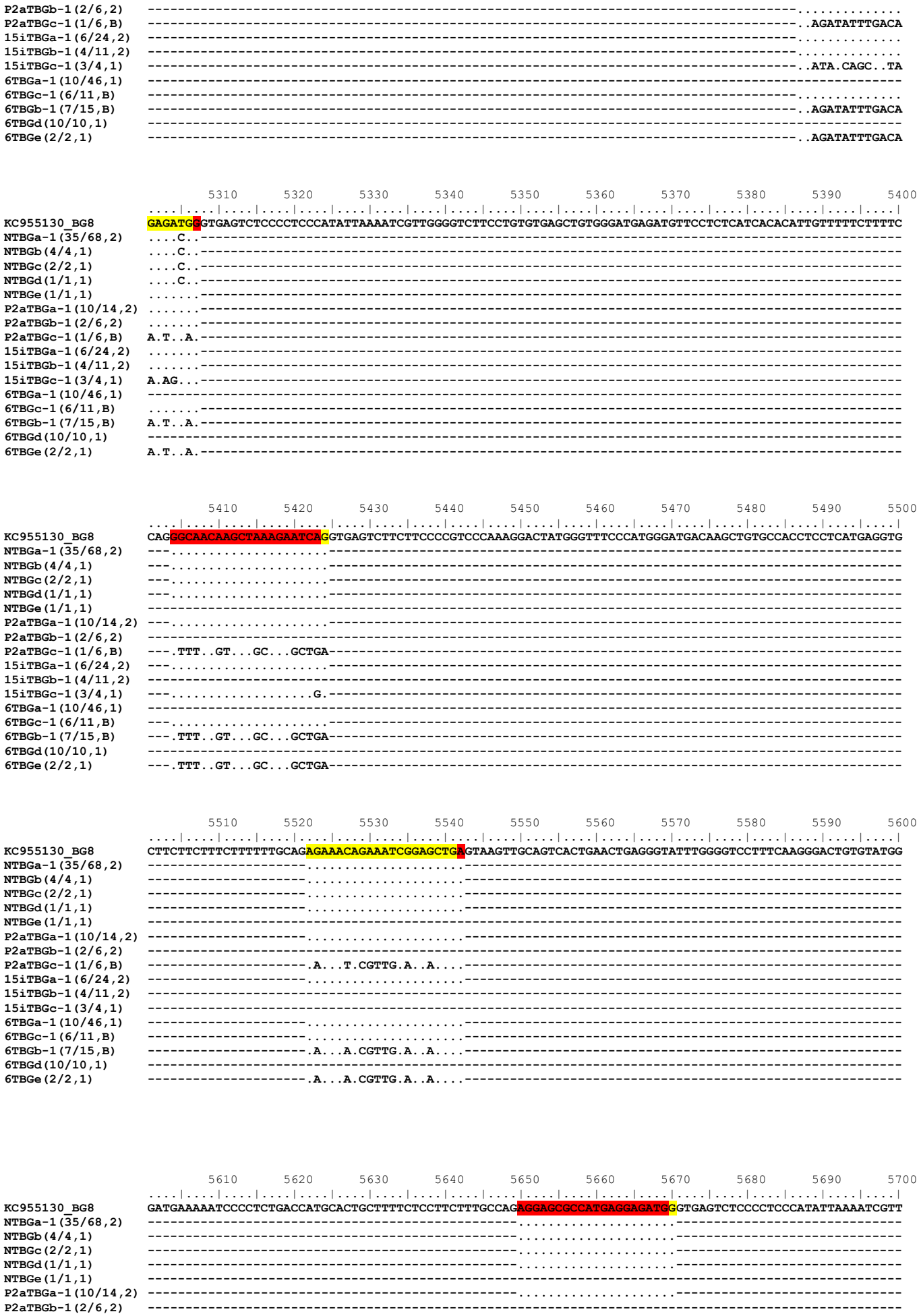

NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
3310 3320 3330 3340 3350 3360 3370 3380 3390 3400
KC955130_BG8
NTBGa-1(35/68,2)
3410 3420 3430 3440 3450 3460 3470 3480 3490 3500
3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
3610 3620 3630 3640 3650 3660 3670 3680 3690 3700

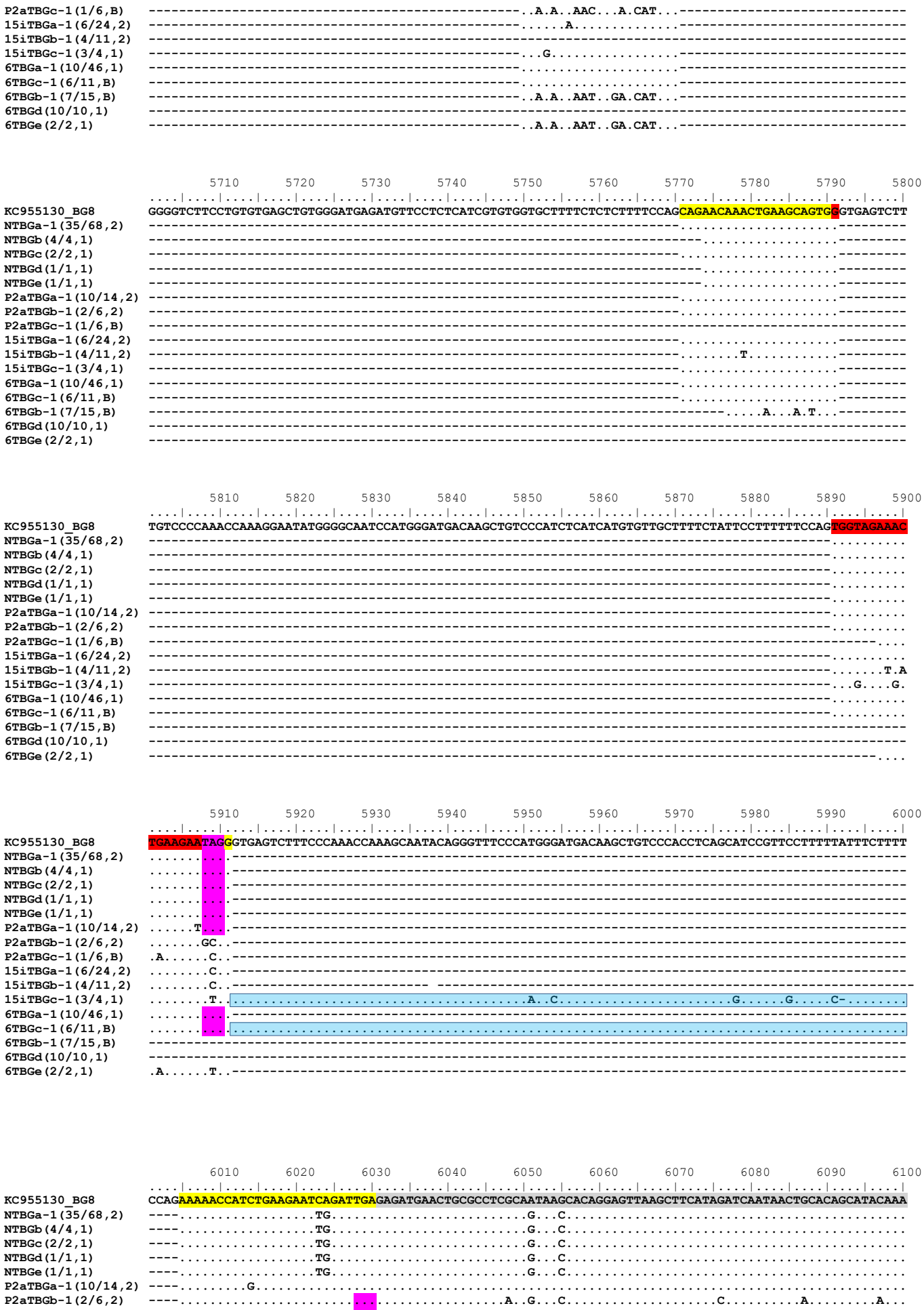

NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
3710 3720 3730 3740 3750 3760 3770 3780 3790 3800
KC955130_BG8 ATTAAATAAAAAATGGGGTCTGCCTGTGTGAGCTGTGAGATGAGATGTTCCTCTCATCATGCGCTGCTTTTCTCTTCCTTTTCCAGAACATCAAACTAAA
3810 3820 3830 3840 3850 3860 3870 3880 3890 3900
KC955130_BG8 GAATTGGGTGAGTCTTCTTTCCCCAACCCCAAGAAATATGCGTTTCCCATGGGATGACAAGCTGTGCCACCTCATCATGCCCTGTTTTTTCTGTCCTTTT
3910 3920 3930 3940 3950 3960 3970 3980 3990 4000
KC955130_BG8 TGCAGAGAAACAGCATTCACAGTTCCGTAAGTTGCAGTCACTGAACTGAAGGAATGTGGGGTCTTCCCAAAGTCCTGCATGTGGGATGAAAAATCCCCTC
4010 4020 4030 4040 4050 4060 4070 4080 4090 4100
KC955130_BG8 TGACCATGCACTGCTTTTCTCCTTCTATTCCA--GAGAGACACTTTCAGAATATGTGTGAGTCTCCCCACCCCTGATAAATAAAAACGTTGGGGTCTTGC

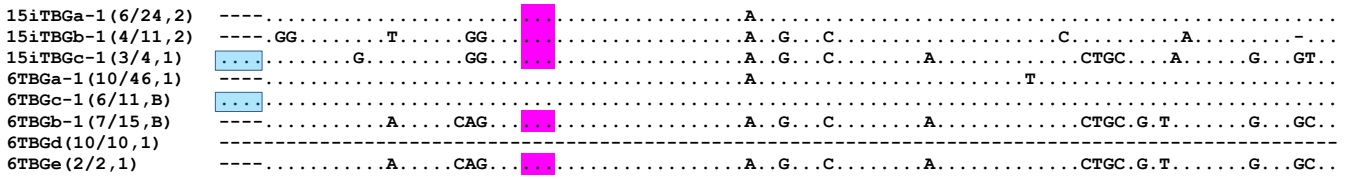

NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
KC955130_BG8
TGTGTGAGCTGTGGGATGAGATGTTCCTCTCATCACACATTGTTTTTCTGATTTTCTTTTGCAGATTTAAGTGCTGGAAAACAGAGTAAGTCTCCCTCCC
4210 4220 4230 4240 4250 4260 4270 4280 4290 4300
KC955130_BG8
TGCACAGAAGGAACTTACGGTTTTCCCATGGGATCAGCCATGGGATCATCATCCGACTCTTCTCATCATGAATTTCGTCTTTCTTTCTTTTGCAGAGAAA
4310 4320 4330 4340 4350 4360 4370 4380 4390 4400
KC955130_BG8
ATGGTTACAAAACTGGGTGAGTCCAACCTCCCAAACTAAATTAAAAACAGTCAGACTTTGTGAGCTGTGGGATGAGACGTTCCTCTCATCATGTGCTGCT
4410 4420 4430 4440 4450 4460 4470 4480 4490 4500
KC955130_BG8
TTCCTTTTACTTTTCCAGAGGAACACTGTGAATGGATGGGTGAGTCTCCCCTCCCAAATTAAAAATGTTGGGGTCTTCCTGTGAGAGCTGTGGGATGAGC

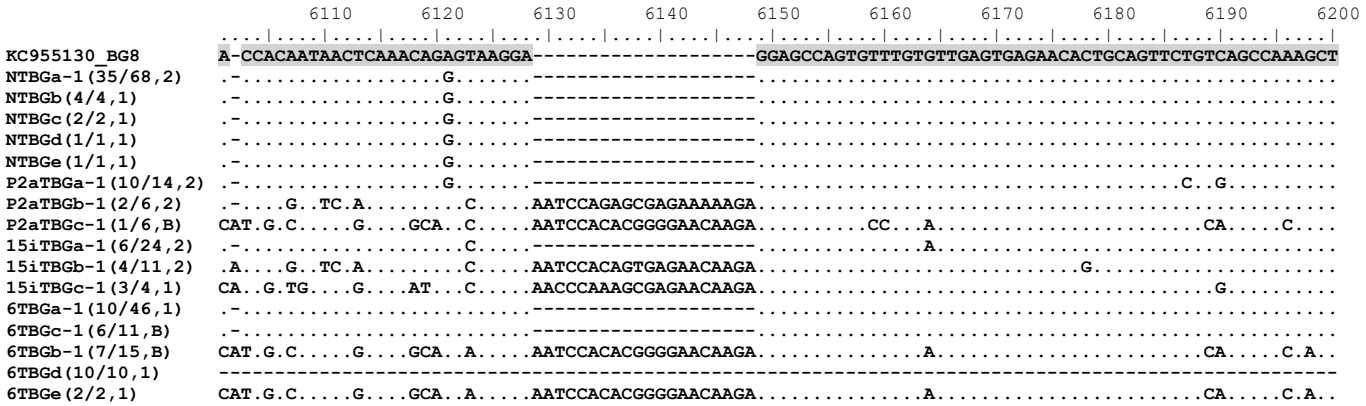

NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
4510 4520 4530 4540 4550 4560 4570 4580 4590 4600
KC955130_BG8 TGTTCCTCTCATCGTGCACTGTTTCTGCTTTTCCTTTGCAGTGAGAAGGAATGTAAAGTTGGGTGAGTCTTCTTCCCCAACCAAAGAGATTCGGAGTCTT
4610 4620 4630 4640 4650 4660 4670 4680 4690 4700
KC955130_BG8 CCATGGGATCAGCCATGGGATGATAACATGAACCTCATCACGTGTTTCTTATTTGTTCCTTTTGCAGAGGCAGCAGCTGTAAAAGTGGGTGAGTCTCCCC
4710 4720 4730 4740 4750 4760 4770 4780 4790 4800
KC955130_BG8 TCCCAAATTAAAAATGTTGGCGTCATCCTGTGAGAGCTGTGGGATGAGCTGTTCCTCTCATCGTGCACTGTTTCTGCTTTTCCTTTGCAGTGAGAAGGAA
4810 4820 4830 4840 4850 4860 4870 4880 4890 4900
KC955130_BG8 TGTAAAGTTGGGTGAGTCTTCTTCCCCAACCAAAGAGATGTGGGGTCTTCCATGGGATCAGCCATGGGATGATAAGCTGAACCTTATCACGTGTTTCTTA

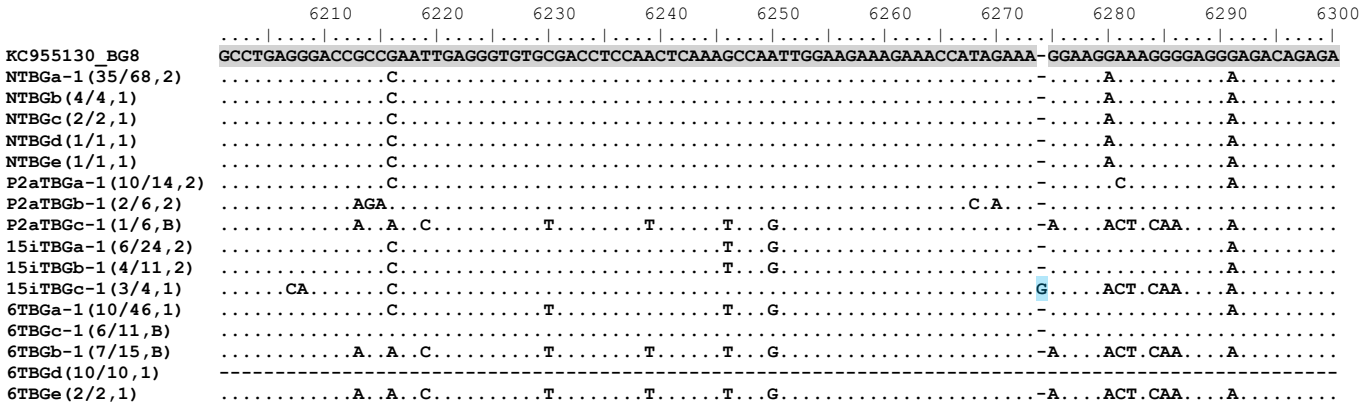

P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
4910 4920 4930 4940 4950 4960 4970 4980 4990 5000
KC955130_BG8 TTTGTTCCTTTTGCAGAGGCAGCAGCTGTAAAAGTGGGTGAGTCCTCCCTCCCAAATCAAATACAAAAGGGGATCTGCCTGTGTGAGCTGTGGGATGAGA
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
5010 5020 5030 5040 5050 5060 5070 5080 5090 5100
KC955130_BG8 TGTTCCTCTCATCACGCATTGTTTTTCTCATTCATTTCCAGGACACAAAGCTAAAGAATCAGGTGAGTCTTCTTCCCTGTCCCAAAGGACTATGGGTTTC
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
5110 5120 5130 5140 5150 5160 5170 5180 5190 5200
KC955130_BG8 CCATGGGATGACAAGCTGTGCCACCTCCTCATGAGGTGCTTCTTCTTTCTTTGTGCAGAGAAACAGAAATCGGAGCTGAGTAAGTTGCAGTCACTGAACT
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
5210 5220 5230 5240 5250 5260 5270 5280 5290 5300
KC955130_BG8 GAGGGAATGTGGGGTCTTCCCAAAGTCCTGCGTATGGGATGAAAAATCCCCTCTGACCATGCACTGCTTTTCTCCTCCTTTGCCAGAGGAGCGCCATGAG
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)

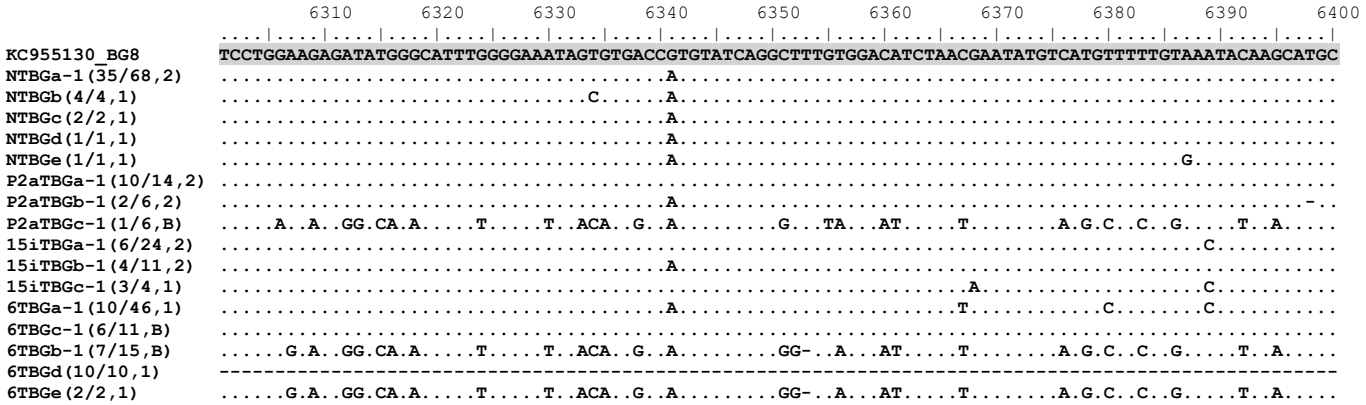

P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
5310 5320 5330 5340 5350 5360 5370 5380 5390 5400
KC955130_BG8 GAGATGGGTGAGTCTCCCCTCCCATATTAAAATCGTTGGGGTCTTCCTGTGTGAGCTGTGGGATGAGATGTTCCTCTCATCACACATTGTTTTTCTTTTC
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
5410 5420 5430 5440 5450 5460 5470 5480 5490 5500
KC955130_BG8 CAGGGCAACAAGCTAAAGAATCAGGTGAGTCTTCTTCCCCGTCCCAAAGGACTATGGGTTTCCCATGGGATGACAAGCTGTGCCACCTCCTCATGAGGTG
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
5510 5520 5530 5540 5550 5560 5570 5580 5590 5600
KC955130_BG8 CTTCTTCTTTCTTTTTTGCAGAGAAACAGAAATCGGAGCTGAGTAAGTTGCAGTCACTGAACTGAGGGTATTTGGGGTCCTTTCAAGGGACTGTGTATGG
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
5610 5620 5630 5640 5650 5660 5670 5680 5690 5700
KC955130_BG8 GATGAAAAATCCCCTCTGACCATGCACTGCTTTTCTCCTTCTTTGCCAGAGGAGCGCCATGAGGAGATGGGTGAGTCTCCCCTCCCATATTAAAATCGTT
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)

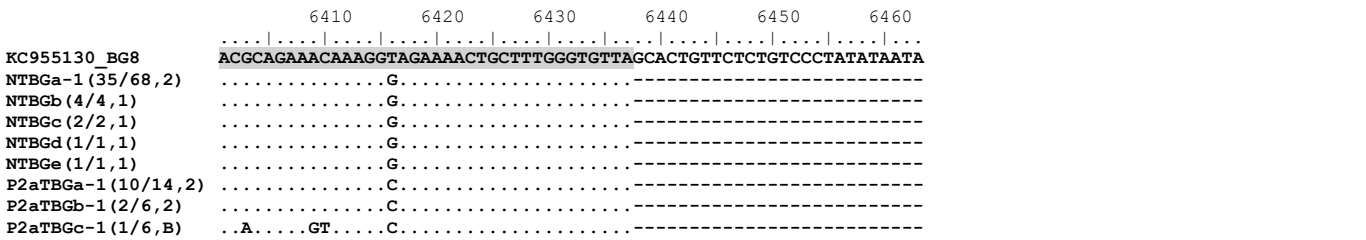

P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
5710 5720 5730 5740 5750 5760 5770 5780 5790 5800
KC955130_BG8 GGGGTCTTCCTGTGTGAGCTGTGGGATGAGATGTTCCTCTCATCGTGTGGTGCTTTTCTCTCTTTTCCAGCAGAACAAACTGAAGCAGTGGGGTGAGTCTT
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
5810 5820 5830 5840 5850 5860 5870 5880 5890 5900
KC955130_BG8 TGTCCCCAAACCAAAGGAATATGGGGCAATCCATGGGATGACAAGCTGTCCCATCTCATCATGTGTTGCTTTTCTATTCCTTTTTTCCAGTGGTAGAAAC
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
5910 5920 5930 5940 5950 5960 5970 5980 5990 6000
KC955130_BG8 TGAAGAATAGGGTGAGTCTTTCCCAAACCAAAGCAATACAGGGTTTCCCATGGGATGACAAGCTGTCCCACCTCAGCATCCGTTCCTTTTTATTTCTTTT
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)
6010 6020 6030 6040 6050 6060 6070 6080 6090 6100
KC955130_BG8 CCAGAAAAACCATCTGAAGAATCAGATTGAGAGATGAACTGCGCCTCGCAATAAGCACAGGAGTTAAGCTTCATAGATCAATAACTGCACAGCATACAAA
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)

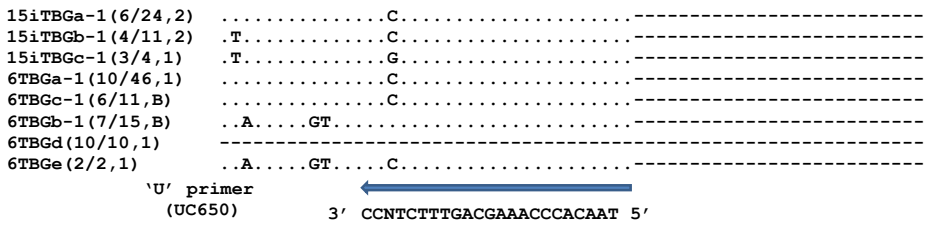

15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)


6110 6120 6130 6140 6150 6160 6170 6180 6190 6200
KC955130_BG8
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)


6210 6220 6230 6240 6250 6260 6270 6280 6290 6300
KC955130_BG8
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)

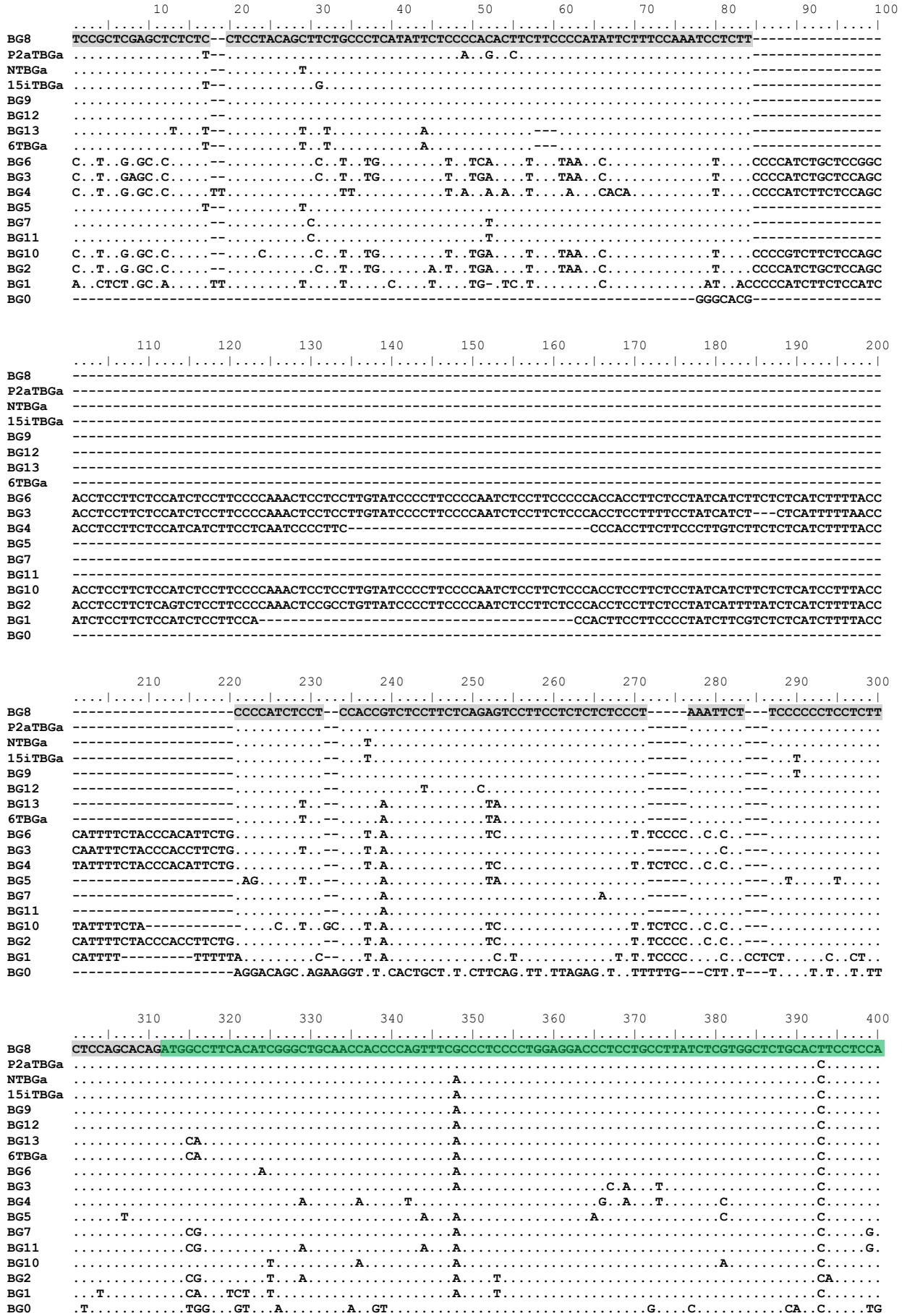

6310 6320 6330 6340 6350 6360 6370 6380 6390 6400
KC955130_BG8
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)
15iTBGa-1(6/24,2)
15iTBGb-1(4/11,2)
15iTBGc-1(3/4,1)
6TBGa-1(10/46,1)
6TBGc-1(6/11,B)
6TBGb-1(7/15,B)
6TBGd(10/10,1)
6TBGe(2/2,1)

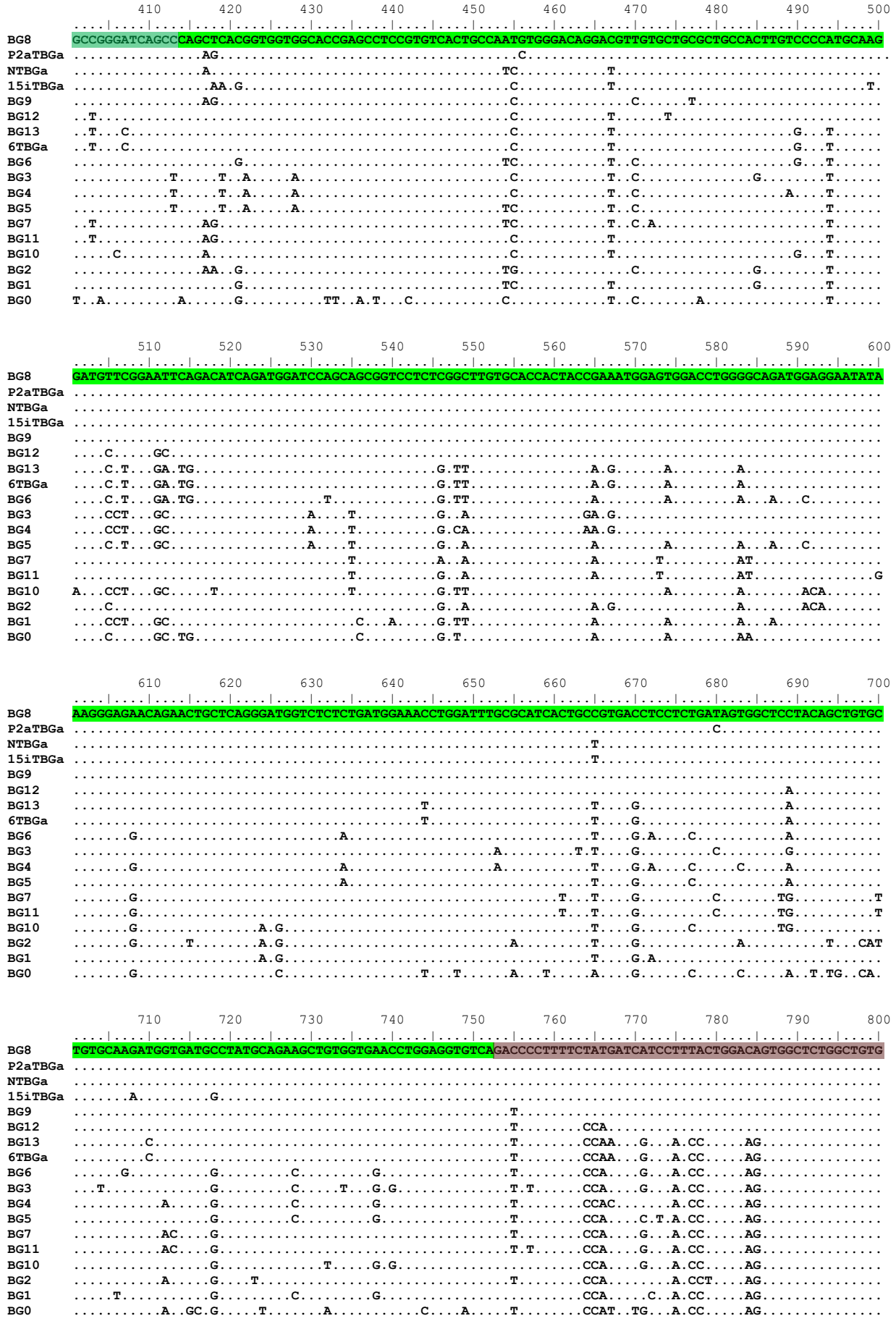

6410 6420 6430 6440 6450 6460
KC955130_BG8
NTBGa-1(35/68,2)
NTBGb(4/4,1)
NTBGc(2/2,1)
NTBGd(1/1,1)
NTBGe(1/1,1)
P2aTBGa-1(10/14,2)
P2aTBGb-1(2/6,2)
P2aTBGc-1(1/6,B)

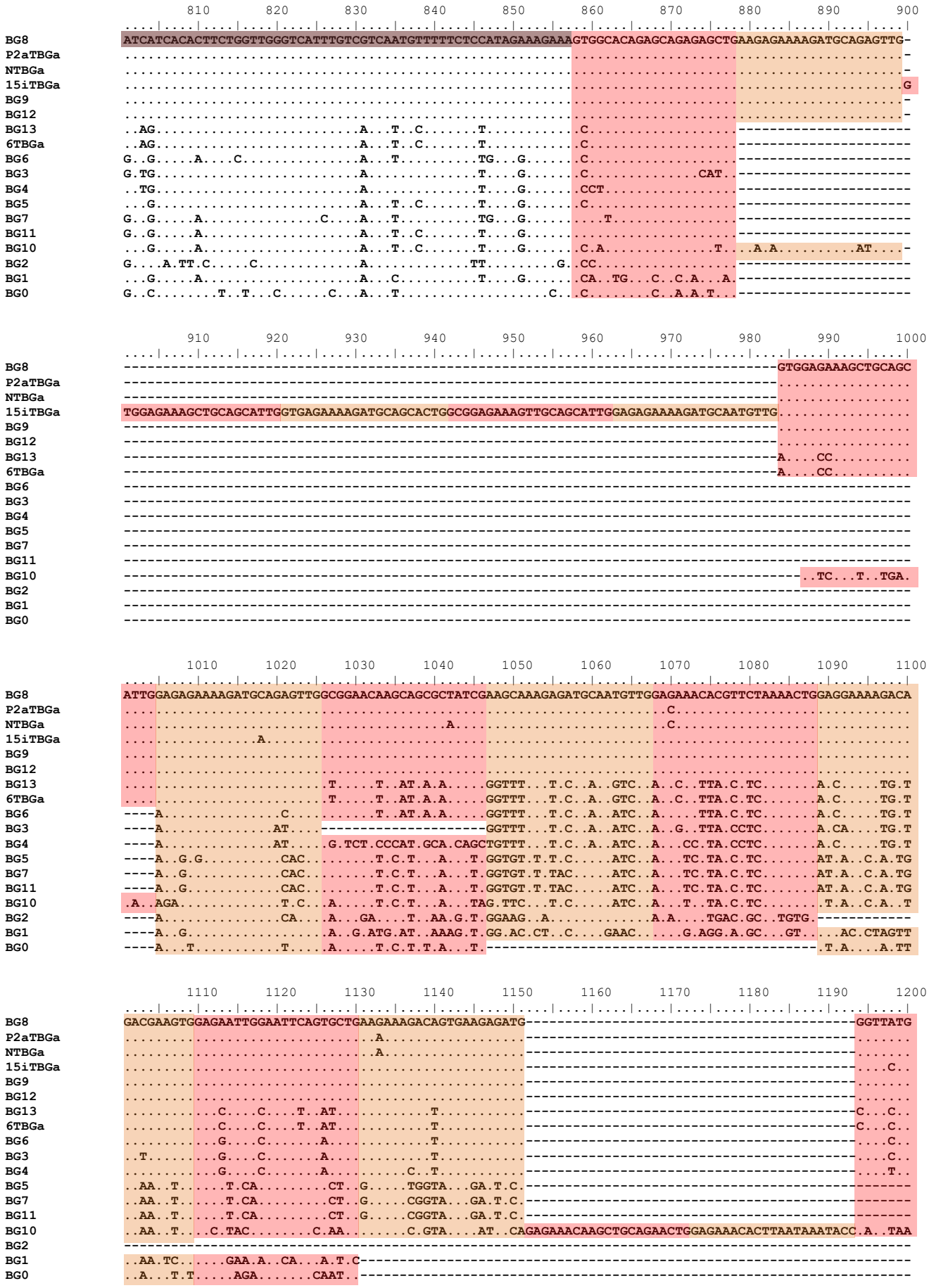


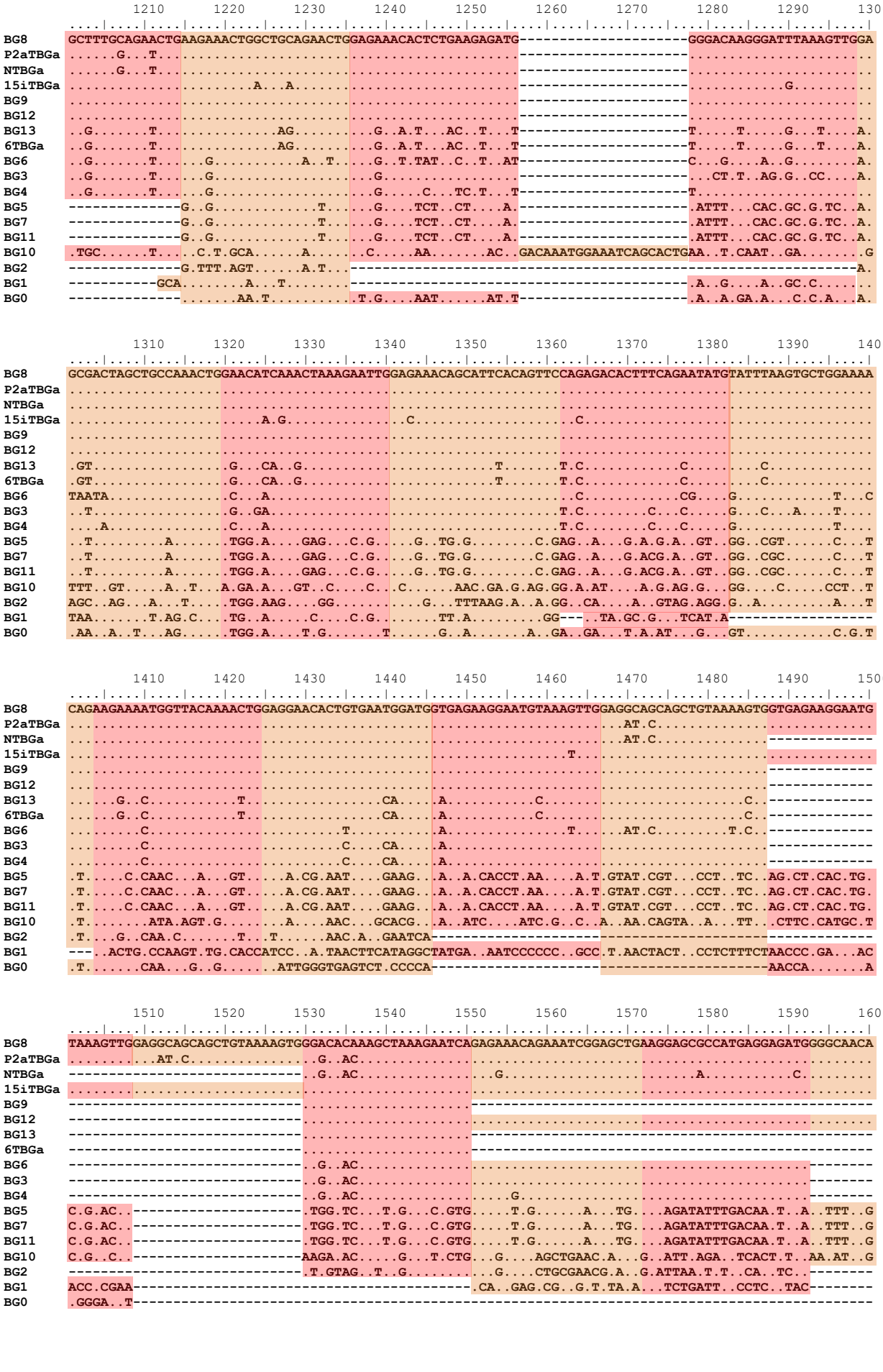
Appendix F. Alignment of the dominantly expressed transcripts for all 16 genes found in T cells from four chicken lines against the BG8 gene (including introns) of B12 haplotype. Names of the transcripts follow the convention: abbreviated line name, "T" for T cells, "BG", the letter "a" representing the most frequently detected clone from the most frequently detected exon 2 sequence (and "b" representing the most frequently detected clone from the second most frequently detected exon 2 sequence, and so forth), a dash and then a number representing the alternative splicing variant with "1" being the most frequently detected clone. Numbers in parenthesis indicate the number of times the depicted sequence was found out of the number of times the gene (based on the exon 2 sequence) was found, followed by the number of independent PCRs in which the sequence was identified (1, found in one PCR; 2 found in 2 PCRs; B, found in one PCR described in this paper and one using B cell cDNA, data not shown). Names of the gene follow the convention: GenBank accession number of B12 BG region genomic sequence, a dash, "BG" and the number of the gene locus for the B12 haplotype. Colours indicate different coding regions: grey, 5'UTR and 3'UTR; dark green, signal sequence; light green, Ig-V domain; brown, transmembrane region; alternating yellow and red, cytoplasmic tail regions (codons from each 18, 21 or 24 nucleotide repeat, with nucleotides in the split codon indicated by the color of the exon in which the majority of the codon is located); purple, in-frame stop codon; light blue, introns (including two positions that are probably nucleotide misincorporations during PCR reaction) which are all deleted in the analyses for "(nearly) full-length conceptual transcripts" (that is, exons without introns). Letters indicate nucleotides, dot indicates identity with BG8 sequence; dash indicates not present in all sequences. Arrows indicate primers with sequences (so that sequence differences in these positions are not necessarily real).

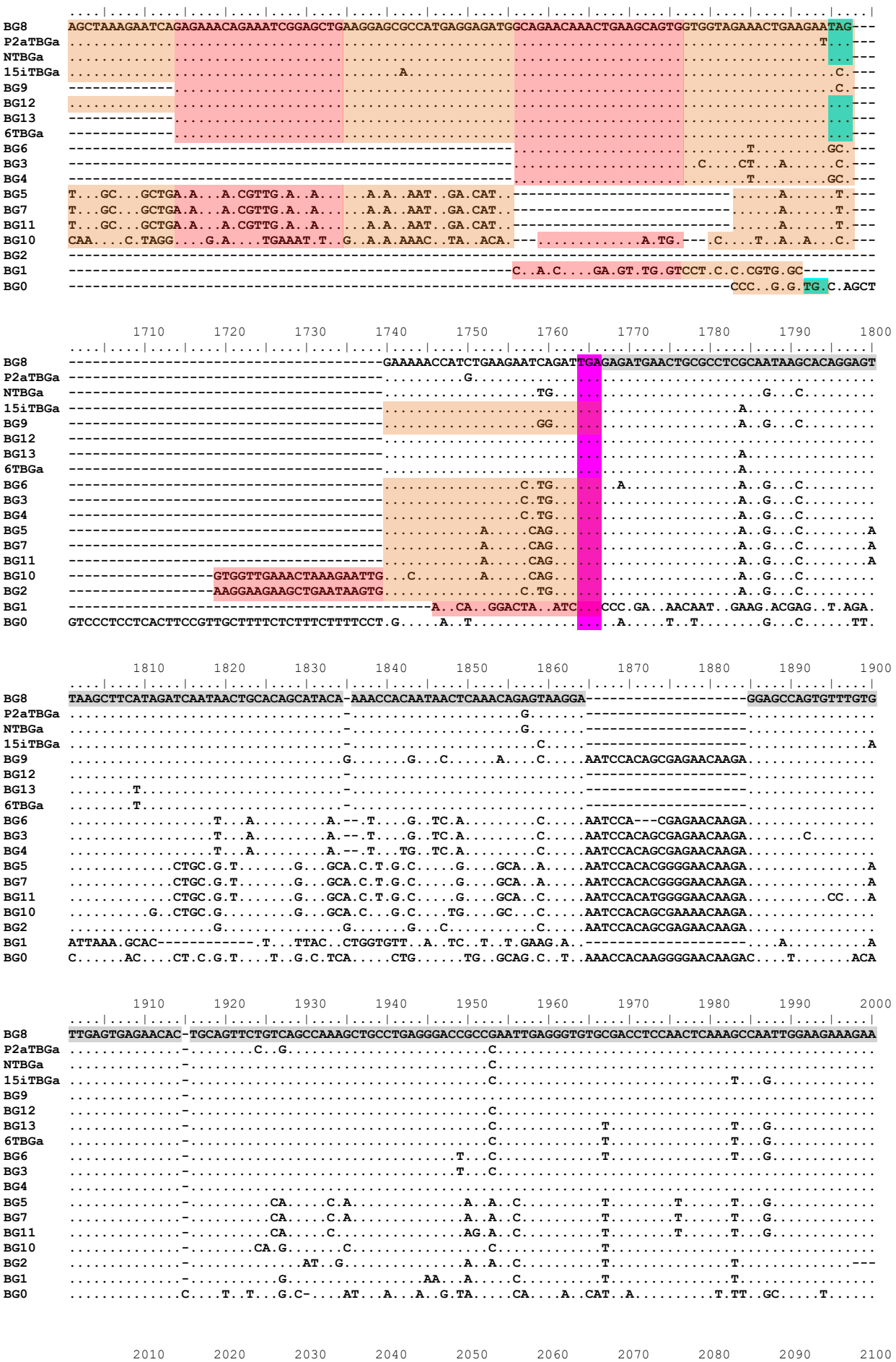
Appendix G.

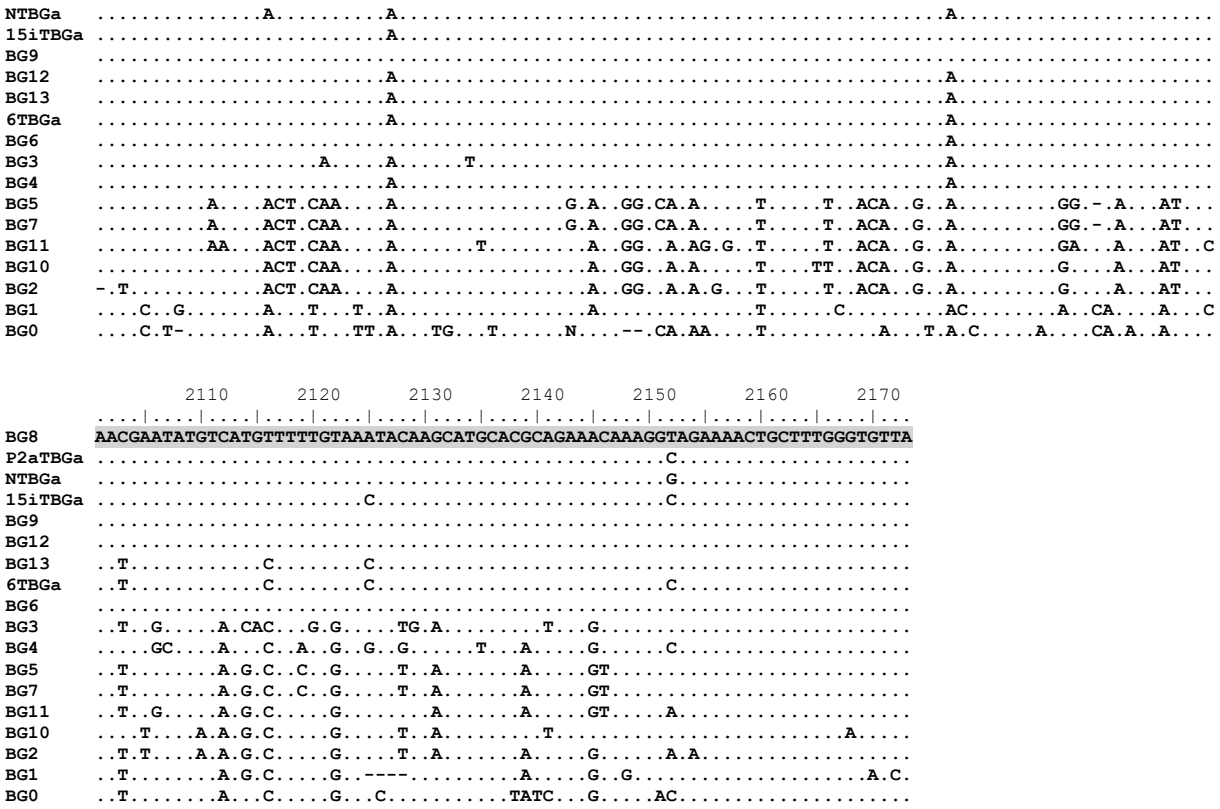

BG8 TCCGCTCGAGCTCTCTC--CTCCTACAGCTTCTGCCCTCATATTCTCCCCACACTTCTTCCCCATATTCTTTCCAAATCCTCTT----------
P2aTBGa .....T--.....A..G..C.....----------
NTBGa .....--.....T.....----------
15iTBGa .....T--.....G.....----------
BG9 .....--.....----------
BG12 .....--.....----------
BG13 .....T...T--.....T..T.....A.....---.....----------
6TBGa .....T--.....T..T.....A.....---.....----------
BG6 C..T..G.GC.C.....--.....C..T..TG.....T..TCA....T...TAA..C.....T....CCCCATCTGCTCCGGC
BG3 C..T..GAGC.C.....--.....C..T..TG.....T..TGA....T...TAA..C.....T....CCCCATCTGCTCCAGC
BG4 C..T..G.GC.C.....TT.....TT.....T.A..A.A..T....A...CACA.....T....CCCCATCTTCTCCAGC
BG5 .....T--.....T.....----------
BG7 .....--.....C.....T.....----------
BG11 .....--.....C.....T.....----------
BG10 C..T..G.GC.C.....--....C.....C..T..TG.....T..TGA....T...TAA..C.....T....CCCCGTCTTCTCCAGC
BG2 C..T..G.GC.C.....--.....C..T..TG.....A.T..TGA....T...TAA..C.....T....CCCCATCTGCTCCAGC
BG1 A..CTCT.GC.A.....TT.....T....T.....C....T....TG-.TC.T.....C.....AT..ACCCCCATCTTCTCCATC
BG0 ----------GGGCACG----------
BG8 ----------
P2aTBGa ----------
NTBGa ----------
15iTBGa ----------
BG9 ----------
BG12 ----------
BG13 ----------
6TBGa ----------
BG6 ACCTCCTTCTCCATCTCCTTCCCCAAACTCCTCCTTGTATCCCCTTCCCCAATCTCCTTCCCCCACCACCTTCTCCTATCATCTTCTCTCATCTTTTACC
BG3 ACCTCCTTCTCCATCTCCTTCCCCAAACTCCTCCTTGTATCCCCTTCCCCAATCTCCTTCTCCCACCTCCTTTTCCTATCATCT---CTCATTTTTAACC
BG4 ACCTCCTTCTCCATCATCTTCCTCAATCCCCTTC----------CCCACCTTCTTCCCTTGTCTTCTCTCATCTTTTACC
BG5 ----------
BG7 ----------
BG11 ----------
BG10 ACCTCCTTCTCCATCTCCTTCCCCAAACTCCTCCTTGTATCCCCTTCCCCAATCTCCTTCTCCCACCTCCTTCTCCTATCATCTTCTCTCATCCTTTACC
BG2 ACCTCCTTCTCAGTCTCCTTCCCCAAACTCCGCCTGTTATCCCCTTCCCCAATCTCCTTCTCCCACCTCCTTCTCCTATCATTTTATCTCATCTTTTACC
BG1 ATCTCCTTCTCCATCTCCTTCCA----------CCACTTCCTTCCCCTATCTTCGTCTCTCATCTTTTACC
BG0 ----------
BG8 ----------CCCCATCTCCT--CCACCGTCTCCTTCTCAGAGTCCTTCCTCTCTCTCCCT-----AAATTCT---TCCCCCCTCCTCTT
P2aTBGa ----------.....--.....-----.....---.....
NTBGa ----------.....--...T.....-----.....---.....
15iTBGa ----------.....--...T.....-----.....---...T.....
BG9 ----------.....--.....-----.....---...T.....
BG12 ----------.....--.....T.....C.....-----.....---.....
BG13 ----------.....T..--.....A.....TA.....-----.....---.....
6TBGa ----------.....T..--.....A.....TA.....-----.....---.....
BG6 CATTTTCTACCCACATTCTG.....--...T.A.....TC.....T.TCCCC..C.C..---.....
BG3 CAATTTCTACCCACCTTCTG.....T..--...T.A.....-----....C..---.....
BG4 TATTTTCTACCCACATTCTG.....--...T.A.....TC.....T.TCTCC..C.C..---.....
BG5 ----------.AG.....T..--.....A.....TA.....-----.....---..T.....T.....
BG7 ----------.....--.....A.....A.....-----.....---.....
BG11 ----------.....--.....A.....-----.....---.....
BG10 TATTTTCTA----------....C..T..GC...T.A.....TC.....T.TCTCC..C.C..---.....
BG2 CATTTTCTACCCACCTTCTG.....--...T.A.....TC.....T.TCCCC..C.C..---.....
BG1 CATTTT---------TTTTTA.....C--...T.A.....C.T.....T.T.TCCCC....C..CCTCT.....C..CT..
BG0 ----------AGGACAGC.AGAAGGT.T.CACTGCT.T.CTTCAG.TT.TTAGAG.T..TTTTTG---CTT.T---T....T.T..T.TT
BG8 CTCCAGCACAGATGGCCTTCACATCGGGCTGCAACCACCCCAGTTTCGCCCTCCCCTGGAGGACCCTCCTGCCTTATCTCGTGGCTCTGCACTTCCTCCA
P2aTBGa .....C.....
NTBGa .....A.....C.....
15iTBGa .....A.....C.....
BG9 .....A.....C.....
BG12 .....A.....C.....
BG13 .....CA.....A.....C.....
6TBGa .....CA.....A.....C.....
BG6 .....A.....A.....C.....
BG3 .....A.....C.A...T.....C.....
BG4 .....A.....A.....T.....G..A...T.....C.....C.....
BG5 .....T.....A...A.....A.....C.....C.....
BG7 .....CG.....A.....C.....G.
BG11 .....CG.....A.....A...A.....C.....G.
BG10 .....T.....A.....A.....A.....C.....
BG2 .....CG.....T...A.....A....T.....CA.....
BG1 ...T.....CA...TCT..T.....A....T.....C.....
BG0 .T.....TGG...GT...A.....A..GT.....G....C.....CA..C.....TG


410 420 430 440 450 460 470 480 490 500
BG8 GCCGGGATCAGCCCAGCTCACGGTGGTGGCACCGAGCCTCCGTGTCACTGCCAATGTGGGACAGGACGTTGTGCTGCGCTGCCACTTGTCCCCATGCAAG
P2aTBGa
NTBGa
15iTBGa
BG9
BG12
BG13
6TBGa
BG6
BG3
BG4
BG5
BG7
BG11
BG10
BG2
BG1
BG0
510 520 530 540 550 560 570 580 590 600
BG8 GATGTTCGGAATTCAGACATCAGATGGATCCAGCAGCGGTCCTCTCGGCTTGTGCACCACTACCGAAATGGAGTGGACCTGGGGCAGATGGAGGAATATA
P2aTBGa
NTBGa
15iTBGa
BG9
BG12
BG13
6TBGa
BG6
BG3
BG4
BG5
BG7
BG11
BG10
BG2
BG1
BG0
610 620 630 640 650 660 670 680 690 700
BG8 AAGGGAGAACAGAACTGCTCAGGGATGGTCTCTCTGATGGAAACCTGGATTTGCGCATCACTGCCGTGACCTCCTCTGATAGTGGCTCCTACAGCTGTGC
P2aTBGa
NTBGa
15iTBGa
BG9
BG12
BG13
6TBGa
BG6
BG3
BG4
BG5
BG7
BG11
BG10
BG2
BG1
BG0
710 720 730 740 750 760 770 780 790 800
BG8 TGTGCAAGATGGTGATGCCTATGCAGAAGCTGTGGTGAACCTGGAGGTGTCAGACCCCTTTTCTATGATCATCCTTTACTGGACAGTGGCTCTGGCTGTG
P2aTBGa
NTBGa
15iTBGa
BG9
BG12
BG13
6TBGa
BG6
BG3
BG4
BG5
BG7
BG11
BG10
BG2
BG1
BG0


BG8
P2aTBGa
NTBGa
15iTBGa
BG9
BG12
BG13
6TBGa
BG6
BG3
BG4
BG5
BG7
BG11
BG10
BG2
BG1
BG0


1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
BG8 GCTTTGCAGAACTGAAGAAACTGGCTGCAGAACTGGAGAAACACTCTGAAGAGATG----------GGGACAAGGGATTTAAAGTTGGA
P2aTBGa ......G...T..----------.....
NTBGa ......G...T..----------.....
15iTBGa .....A...A..----------.....G.....
BG9 ----------.....
BG12 ----------.....
BG13 ..G.....T.....AG.....G..A.T...AC..T...T----------T.....T.....G...T....A.
6TBGa ..G.....T.....AG.....G..A.T...AC..T...T----------T.....T.....G...T....A.
BG6 ..G.....T.....G.....A..T.....G..T.TAT..C..T..AT----------C...G....A..G.....A.
BG3 ..G.....T.....G.....G.....----------...CT.T..AG.G..CC....A.
BG4 ..G.....T.....G.....G.....C...TC.T...T----------T.....G
BG5 ----------G..G.....T.....G....TCT..CT....A.----------.ATTT...CAC.GC.G.TC..A.
BG7 ----------G..G.....T.....G....TCT..CT....A.----------.ATTT...CAC.GC.G.TC..A.
BG11 ----------G..G.....T.....G....TCT..CT....A.----------.ATTT...CAC.GC.G.TC..A.
BG10 .TGC.....T....C.T.GCA.....A.....C.....AA.....AC..GACAAATGGAAATCAGCACTGAA..T.CAAT..GA.....G
BG2 ----------G.TTT.AGT.....A.T...----------A.
BG1 ----------GCA.....A...T.....----------.A..G....A..GC.C.....
BG0 ----------.....AA.T.....T.G....AAT.....AT.T----------.A..A.GA.A...C.C.A...A.
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
BG8 GCGACTAGCTGCCAAACTGGAACATCAAACTAAAGAATTGGAGAAACAGCATTCACAGTTCCAGAGACACTTTCAGAATATGTATTTAAGTGCTGGAAAA
P2aTBGa .....
NTBGa .....
15iTBGa .....A.G.....C.....C.....
BG9 .....
BG12 .....
BG13 .GT.....G...CA..G.....T.....T.C.....C.....C.....
6TBGa .GT.....G...CA..G.....T.....T.C.....C.....C.....
BG6 TAATA.....C...A.....C.....CG....G.....T...C
BG3 ..T.....G..GA.....T.C.....C...C.....G...C...A....T....
BG4 ....A.....C...A.....T.C.....C...C.....G.....T....
BG5 ..T.....A.....TGG.A....GAG...C.G....G..TG.G.....C.GAG..A...G.A.G.A..GT..GG..CGT.....C...T
BG7 ..T.....A.....TGG.A....GAG...C.G....G..TG.G.....C.GAG..A...G.ACG.A..GT..GG..CGC.....C...T
BG11 ..T.....A.....TGG.A....GAG...C.G....G..TG.G.....C.GAG..A...G.ACG.A..GT..GG..CGC.....C...T
BG10 TTT..GT.....A..T...A.GA.A...GT..C....C...C.....AAC.GA.G.AG.GG.A.AT....A.G.AG.G...GG....C.....CCT..T
BG2 AGC..AG...A...T.....TGG.AAG....GG.....G...TTTAAG.A..A.GG..CA....A..GTAG.AGG.G..A.....A...T
BG1 TAA.....T.AG.C....TG..A....C....C.G.....TT.A.....GG---..TA.GC.G...TCAT.A----------
BG0 .AA..A..T...AG.....TGG.A....T.G.....T.....G..A.....A..GA..GA...T.A.AT...G...GT.....C.G.T
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
BG8 CAGAAGAAAATGGTTACAAAACTGGAGGAACACTGTGAATGGATGGTGAGAAGGAATGTAAAGTTGGAGGCAGCAGCTGTAAAAGTGGTGAGAAGGAATG
P2aTBGa .....AT.C.....
NTBGa .....AT.C.....----------
15iTBGa .....T.....
BG9 .....----------
BG12 .....----------
BG13 .....G..C.....T.....CA....A.....C.....C..----------
6TBGa .....G..C.....T.....CA....A.....C.....C..----------
BG6 .....C.....T.....A.....T.....AT.C.....T.C..----------
BG3 .....C.....C....CA....A.....----------
BG4 .....C.....C....CA....A.....----------
BG5 .T.....C.CAAC...A...GT.....A.CG.AAT....GAAG...A..A.CACCT.AA....A.T.GTAT.CGT...CCT..TC..AG.CT.CAC.TG.
BG7 .T.....C.CAAC...A...GT.....A.CG.AAT....GAAG...A..A.CACCT.AA....A.T.GTAT.CGT...CCT..TC..AG.CT.CAC.TG.
BG11 .T.....C.CAAC...A...GT.....A.CG.AAT....GAAG...A..A.CACCT.AA....A.T.GTAT.CGT...CCT..TC..AG.CT.CAC.TG.
BG10 .T.....ATA.AGT.G.....A....AAC...GCACG...A..ATC....ATC.G..C..A..AA.CAGTA..A...TT..CTTC.CATGC.T
BG2 .T....G..CAA.C.....T...T.....AAC.A..GAATCA----------
BG1 ---..ACTG.CCAAGT.TG.CACCATCC..A.TAACTTCATAGGCTATGA..AATCCCCCCC..GCC.T.AACTACT..CCTCTTTCTAACCC.GA...AC
BG0 .T.....CAA...G..G.....ATTGGGTGAGTCT.CCCCA----------AACCA.....A
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
BG8 TAAAGTTGGAGGCAGCAGCTGTAAAAGTGGGACACAAAGCTAAAGAATCAGAGAAACAGAAATCGGAGCTGAAGGAGCGCCATGAGGAGATGGGGCAACA
P2aTBGa .....AT.C.....G..AC.....
NTBGa ----------G..AC.....G.....A.....C.....
15iTBGa .....
BG9 ----------
BG12 ----------.....
BG13 ----------
6TBGa ----------
BG6 ----------G..AC.....--------
BG3 ----------G..AC.....--------
BG4 ----------G..AC.....G.....--------
BG5 C.G.AC..----------.TGG.TC...T.G...C.GTG.....T.G.....A...TG...AGATATTTGACAA.T..A..TTT..G
BG7 C.G.AC..----------.TGG.TC...T.G...C.GTG.....T.G.....A...TG...AGATATTTGACAA.T..A..TTT..G
BG11 C.G.AC..----------.TGG.TC...T.G...C.GTG.....T.G.....A...TG...AGATATTTGACAA.T..A..TTT..G
BG10 C.G..C..----------AAGA.AC.....G...T.CTG...G....AGCTGAAC.A...G..ATT.AGA..TCACT.T..AA.AT..G
BG2 ----------.T.GTAG..T..G.....G....CTGCGAACG.A..G.ATTAA.T.T..CA..TC..--------
BG1 ACC.CGAA----------.CA..GAG.CG..G.T.TA.A...TCTGATT..CCTC..TAC--------
BG0 .GGGA..T----------
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700


BG8
P2aTBGa
NTBGa
15iTBGa
BG9
BG12
BG13
6TBGa
BG6
BG3
BG4
BG5
BG7
BG11
BG10
BG2
BG1
BG0
1710
1720
1730
1740
1750
1760
1770
1780
1790
1800
1810
1820
1830
1840
1850
1860
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
2030
2040
2050
2060
2070
2080
2090
2100



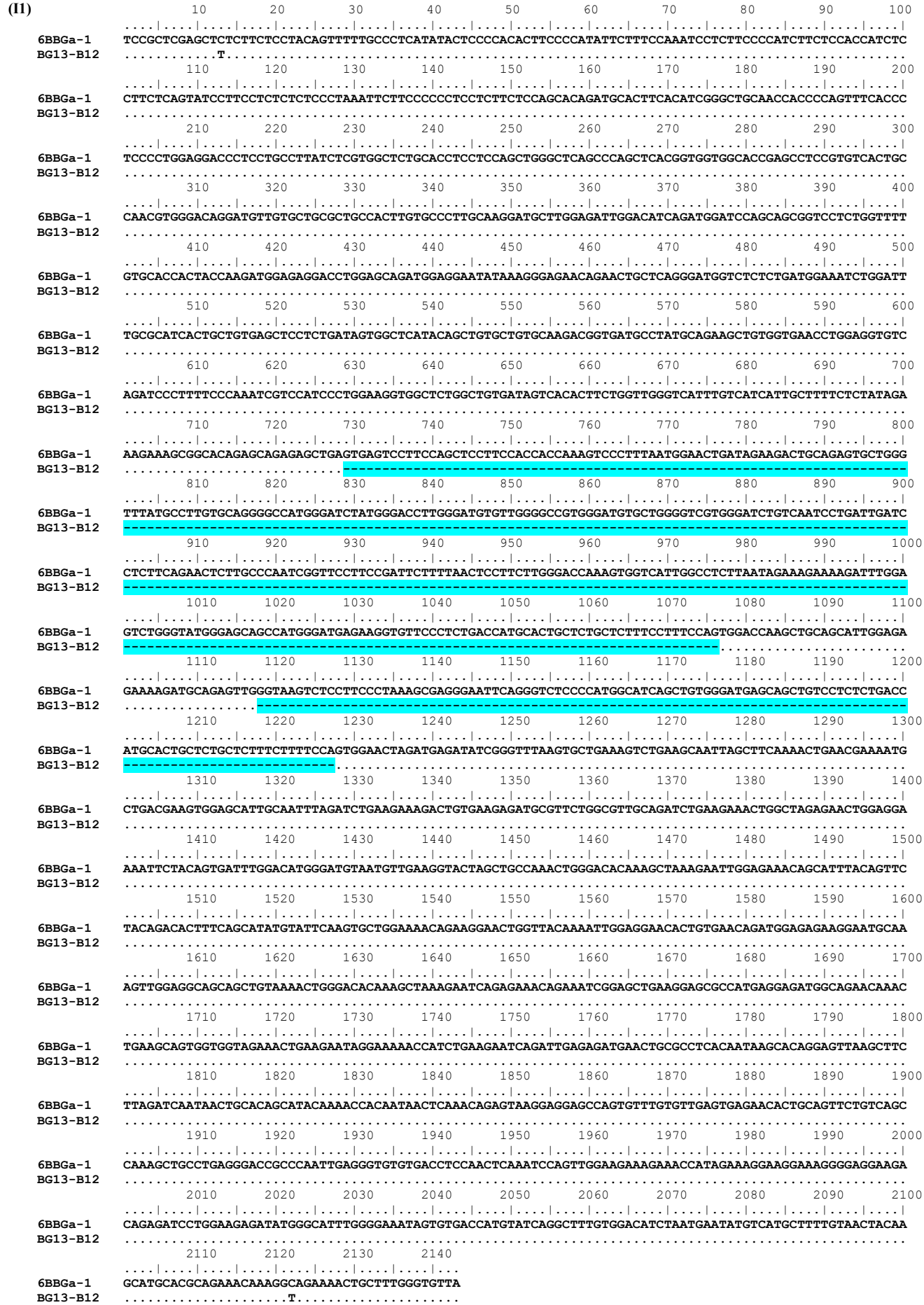
Appendix G. Alignments of nucleotide sequences from the full-length conceptual transcripts for the dominantly expressed genes from T cells of four chicken lines, and for the 14 BG genes of the B12 haplotype. Names of the transcripts follow the convention: abbreviated line name, "T" for T cells, "BG" and the letter "a" representing the most frequently detected clone from the most frequently detected exon 2 sequence. Names of the genes follow the convention "BG" and the number of the gene locus for the B12 haplotype. Colors indicate different coding regions (grey, 5'UTR and 3'UTR; dark green, signal sequence; light green, Ig-V domain; brown, transmembrane region; alternating yellow and red, cytoplasmic tail regions (codons from each 18, 21 or 24 nucleotide repeat); purple, in-frame stop codon; light blue, introns (including two positions that are probably nucleotide misincorporations during PCR reaction) which are all deleted in the analyses for "(nearly) full-length conceptual transcripts" (that is, exons without introns). In this figure, the colors of the nucleotides of split codons reflect the protein region rather than the exon (for instance, last amino acid of the signal sequence is assigned to the signal sequence even though the last two nucleotides of the codon are part of exon 2 that encodes the Ig-V domain). Letters indicate nucleotides, dots indicate identity with BG8 sequence; dashes indicate no sequence present compared to one or more of the other sequences.

Appendix H. Sequences deposited in GenBank with accession numbers

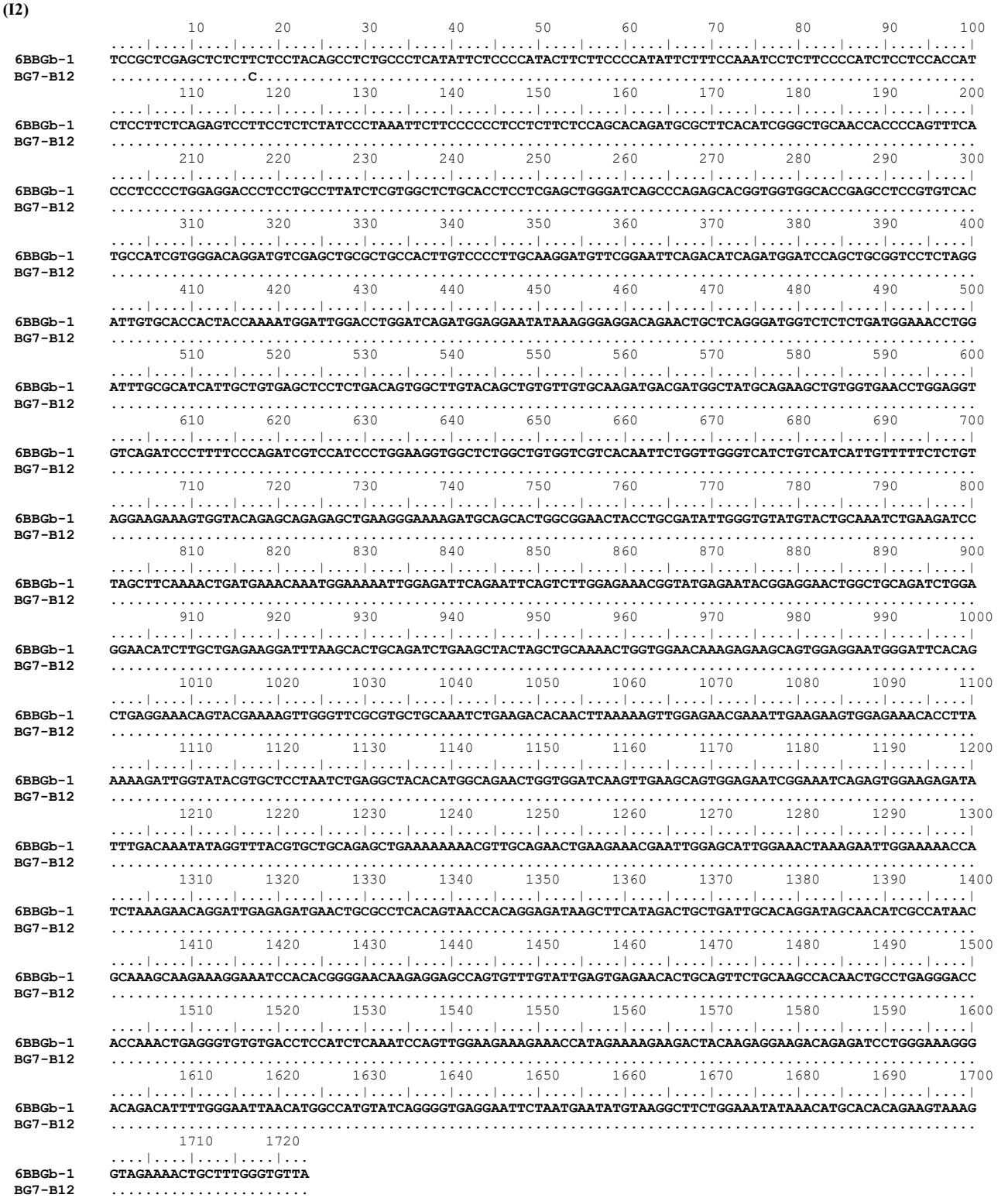
Sequence name	GenBank accession number
NTBGa-1	MH156615
NTBGa-3	MH156616
NTBGa-4	MH156617
NTBGa-16	MH156618
NTBGb	MH156619
NTBGc	MH156620
NTBGd	MH156621
NTBGe	MH156622
P2aTBGa-1	MH156623
P2aTBGb-1	MH156624
P2aTBGc-1	MH156625
P2aTBGc-5	MH156626
15iTBGa-1	MH156627
15iTBGa-2	MH156628
15iTBGa-3	MH156629
15iTBGa-4	MH156630
15iTBGa-5	MH156631
15iTBGa-7	MH156632
15iTBGb-1	MH156633
15iTBGb-2	MH156634
15iTBGb-5	MH156635
15iTBGc-1	MH156636
6TBGa-1	MH156637
6TBGa-2	MH156638
6TBGb-1	MH156639
6TBGb-2	MH156640
6TBGc-1	MH156641
6TBGd	MH156642
6TBGe	MH156643

Appendix I.

(I1)



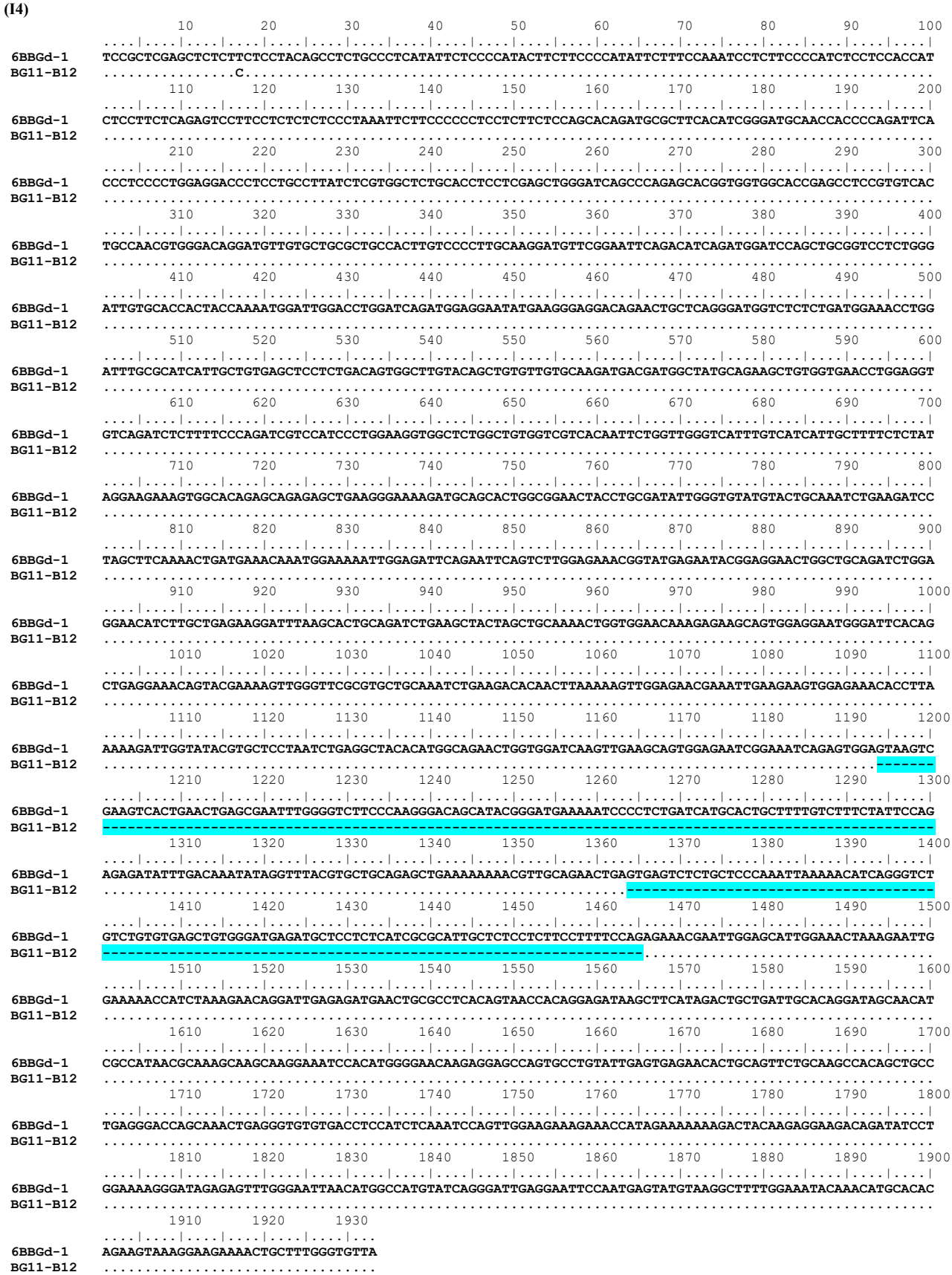
(I2)


10 20 30 40 50 60 70 80 90 100
6BBGb-1 TCCGCTCGAGCTCTCTTCTCCTACAGCCTCTGCCCTCATATTCTCCCCATACTTCTTCCCCATATTCTTTCCAAATCCTCTTCCCCATCTCCTCCACCAT
BG7-B12 .....C.....
110 120 130 140 150 160 170 180 190 200
6BBGb-1 CTCCTTCTCAGAGTCCTTCCTCTCTATCCCTAAATTCTTCCCCCCTCCTCTTCTCCAGCACAGATGCGCTTCACATCGGGCTGCAACCACCCCAGTTTCA
BG7-B12 .....
210 220 230 240 250 260 270 280 290 300
6BBGb-1 CCCTCCCCTGGAGGACCCTCCTGCCTTATCTCGTGGCTCTGCACCTCCTCGAGCTGGGATCAGCCCAGAGCACGGTGGTGGCACCGAGCCTCCGTGTCAC
BG7-B12 .....
310 320 330 340 350 360 370 380 390 400
6BBGb-1 TGCCATCGTGGGACAGGATGTCGAGCTGCGCTGCCACTTGTCCCCTTGCAAGGATGTTCGGAATTCAGACATCAGATGGATCCAGCTGCGGTCCTCTAGG
BG7-B12 .....
410 420 430 440 450 460 470 480 490 500
6BBGb-1 ATTGTGCACCACTACCAAAATGGATTGGACCTGGATCAGATGGAGGAATATAAAGGGAGGACAGAACTGCTCAGGGATGGTCTCTCTGATGGAAACCTGG
BG7-B12 .....
510 520 530 540 550 560 570 580 590 600
6BBGb-1 ATTTGCGCATCATTGCTGTGAGCTCCTCTGACAGTGGCTTGTACAGCTGTGTTGTGCAAGATGACGATGGCTATGCAGAAGCTGTGGTGAACCTGGAGGT
BG7-B12 .....
610 620 630 640 650 660 670 680 690 700
6BBGb-1 GTCAGATCCCTTTTCCCAGATCGTCCATCCCTGGAAGGTGGCTCTGGCTGTGGTCGTCACAATTCTGGTTGGGTCATCTGTCATCATTGTTTTTCTCTGT
BG7-B12 .....
710 720 730 740 750 760 770 780 790 800
6BBGb-1 AGGAAGAAAGTGGTACAGAGCAGAGAGCTGAAGGGAAAAGATGCAGCACTGGCGGAACTACCTGCGATATTGGGTGTATGTACTGCAAATCTGAAGATCC
BG7-B12 .....
810 820 830 840 850 860 870 880 890 900
6BBGb-1 TAGCTTCAAAACTGATGAAACAAATGGAAAAATTGGAGATTCAGAATTCAGTCTTGGAGAAACGGTATGAGAATACGGAGGAACTGGCTGCAGATCTGGA
BG7-B12 .....
910 920 930 940 950 960 970 980 990 1000
6BBGb-1 GGAACATCTTGCTGAGAAGGATTTAAGCACTGCAGATCTGAAGCTACTAGCTGCAAAACTGGTGGAACAAAGAGAAGCAGTGGAGGAATGGGATTCACAG
BG7-B12 .....
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
6BBGb-1 CTGAGGAAACAGTACGAAAAGTTGGGTTCGCGTGCTGCAAATCTGAAGACACAACTTAAAAAGTTGGAGAACGAAATTGAAGAAGTGGAGAAACACCTTA
BG7-B12 .....
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
6BBGb-1 AAAAGATTGGTATACGTGCTCCTAATCTGAGGCTACACATGGCAGAACTGGTGGATCAAGTTGAAGCAGTGGAGAATCGGAAATCAGAGTGGAAGAGATA
BG7-B12 .....
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
6BBGb-1 TTTGACAAATATAGGTTTACGTGCTGCAGAGCTGAAAAAAAACGTTGCAGAACTGAAGAAACGAATTGGAGCATTGGAAACTAAAGAATTGGAAAAACCA
BG7-B12 .....
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
6BBGb-1 TCTAAAGAACAGGATTGAGAGATGAACTGCGCCTCACAGTAACCACAGGAGATAAGCTTCATAGACTGCTGATTGCACAGGATAGCAACATCGCCATAAC
BG7-B12 .....
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
6BBGb-1 GCAAAGCAAGAAAGGAAATCCACACGGGGAACAAGAGGAGCCAGTGTTTGTATTGAGTGAGAACACTGCAGTTCTGCAAGCCACAACTGCCTGAGGGACC
BG7-B12 .....
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
6BBGb-1 ACCAAACTGAGGGTGTGTGACCTCCATCTCAAATCCAGTTGGAAGAAAGAAACCATAGAAAAGAAGACTACAAGAGGAAGACAGAGATCCTGGGAAAGGG
BG7-B12 .....
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
6BBGb-1 ACAGACATTTTGGGAATTAACATGGCCATGTATCAGGGGTGAGGAATTCTAATGAATATGTAAGGCTTCTGGAAATATAAACATGCACACAGAAGTAAAG
BG7-B12 .....
1710 1720
6BBGb-1 GTAGAAAACTGCTTTGGGTGTTA
BG7-B12 .....

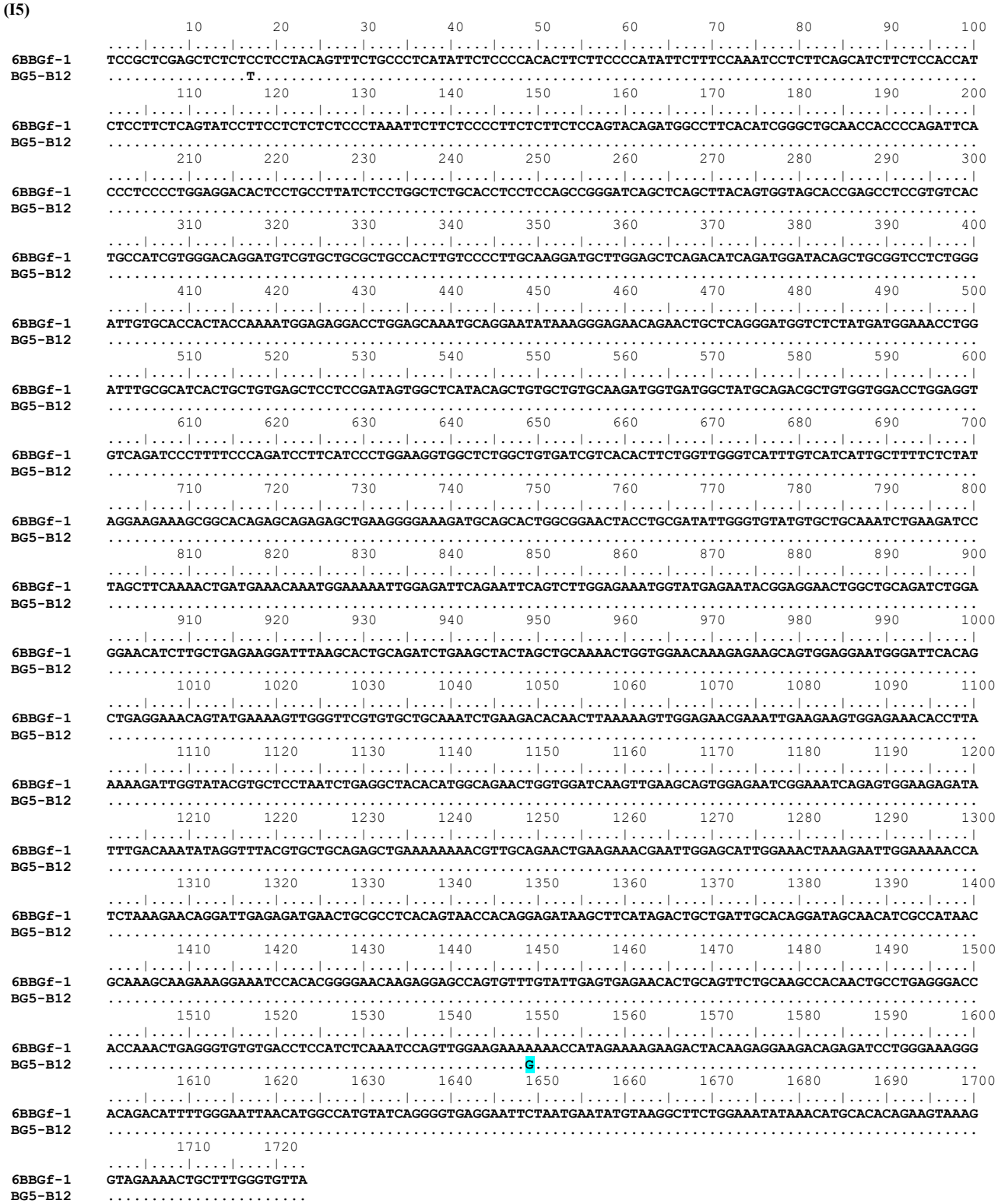
(I3)

```
                10        20        30        40        50        60        70        80        90        100
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6BBGc-1    TCCGCTCGAGCTCTCTTCTCCTACAGCTTCTGCCCTCATATTCTCCCCACACTTCTTCCCCATATTCTTTCCAAATCCTCTTCCCCATCTCCTCCACCGT
BG8-B12    ................C.......................................................................................
                110       120       130       140       150       160       170       180       190       200
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6BBGc-1    CTCCTTCTCAGAGTCCTTCCTCTCTCTCCCTAAATTCTTCCCCCCTCCTCTTCTCCAGCACAGATGGCCTTCACATCGGGCTGCAACCACCCCAGTTTCG
BG8-B12    ................................................................................................
                210       220       230       240       250       260       270       280       290       300
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6BBGc-1    CCCTCCCCTGGAGGACCCTCCTGCCTTATCTCGTGGCTCTGCACTTCCTCCAGCCGGGATCAGCCCAGCTCACGGTGGTGGCACCGAGCCTCCGTGTCAC
BG8-B12    ................................................................................................
                310       320       330       340       350       360       370       380       390       400
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6BBGc-1    TGCCAATGTGGGACAGGACGTTGTGCTGCGCTGCCACTTGTCCCCATGCAAGGATGTTCGGAATTCAGACATCAGATGGATCCAGCAGCGGTCCTCTCGG
BG8-B12    ................................................................................................
                410       420       430       440       450       460       470       480       490       500
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6BBGc-1    CTTGTGCACCACTACCGAAATGGAGTGGACCTGGGGCAGATGGAGGAATATAAAGGGAGAACAGAACTGCTCAGGGATGGTCTCTCTGATGGAAACCTGG
BG8-B12    ................................................................................................
                510       520       530       540       550       560       570       580       590       600
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6BBGc-1    ATTTGCGCATCACTGCCGTGACCTCCTCTGATAGTGGCTCCTACAGCTGTGCTGTGCAAGATGGTGATGCCTATGCAGAAGCTGTGGTGAACCTGGAGGT
BG8-B12    ................................................................................................
                610       620       630       640       650       660       670       680       690       700
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6BBGc-1    GTCAGACCCCTTTTCTATGATCATCCTTTACTGGACAGTGGCTCTGGCTGTGATCATCACACTTCTGGTTGGGTCATTTGTCGTCAATGTTTTTCTCCAT
BG8-B12    ................................................................................................
                710       720       730       740       750       760       770       780       790       800
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6BBGc-1    AGAAAGAAAGTGGCACAGAGCAGAGAGCTGAAGAGAAAAGATGCAGAGTTGGTGGAGAAAGCTGCAGCATTGGAGAGAAAAGATGCAGAGTTGGCGGAAC
BG8-B12    ................................................................................................
                810       820       830       840       850       860       870       880       890       900
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6BBGc-1    AAGCAGCGCTATCGAAGCAAAGAGATGCAATGTTGGAGAAACACGTTCTAAAACTGGAGGAAAAGACAGACGAAGTGGAGAATTGGAATTCAGTGCTGAA
BG8-B12    ................................................................................................
                910       920       930       940       950       960       970       980       990       1000
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6BBGc-1    GAAAGACAGTGAAGAGATGGGTTATGGCTTTGCAGAACTGAAGAAACTGGCTGCAGAACTGGAGAAACACTCTGAAGAGATGGGGACAAGGGATTTAAAG
BG8-B12    ................................................................................................
                1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6BBGc-1    TTGGAGCGACTAGCTGCCAAACTGGAACATCAAACTAAAGAATTGGAGAAACAGCATTCACAGTTCCAGAGACACTTTCAGAATATGTATTTAAGTGCTG
BG8-B12    ................................................................................................
                1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6BBGc-1    GAAAACAGAAGAAAATGGTTACAAAACTGGAGGAACACTGTGAATGGATGGTGAGAAGGAATGTAAAGTTGGAGGCAGCAGCTGTAAAAGTGGTGAGAAG
BG8-B12    ................................................................................................
                1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6BBGc-1    GAATGTAAAGTTGGAGGCAGCAGCTGTAAAAGTGGGACACAAAGCTAAAGAATCAGAGAAACAGAAATCGGAGCTGAAGGAGCGCCATGAGGAGATGGGG
BG8-B12    ................................................................................................
                1310      1320      1330      1340      1350      1360      1370      1380      1390      1400
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6BBGc-1    CAACAAGCTAAAGAATCAGAGAAACAGAAATCGGAGCTGAAGGAGCGCCATGAGGAGATGGCAGAACAAACTGAAGCAGTGGTGGTAGAAACTGAAGAAT
BG8-B12    ................................................................................................
                1410      1420      1430      1440      1450      1460      1470      1480      1490      1500
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6BBGc-1    AGGGTGAGTCTTTCCCAAACCAAAGCAATACAGGGTTTCCCATGGGATGACAAGCTGTCCCACCTCAGCATCCGTTCCTTTTTATTTCTTTTCCAGAAAA
BG8-B12    ...----------------------------------------------------------------------------------------....
                1510      1520      1530      1540      1550      1560      1570      1580      1590      1600
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6BBGc-1    ACCATCTGAAGAATCAGATTGAGAGATGAACTGCGCCTCGCAATAAGCACAGGAGTTAAGCTTCATAGATCAATAACTGCACAGCATACAAAACCACAAT
BG8-B12    ................................................................................................
                1610      1620      1630      1640      1650      1660      1670      1680      1690      1700
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6BBGc-1    AACTCAAACAGAGTAAGGAGGAGCCAGTGTTTGTGTTGAGTGAGAACACTGCAGTTCTGTCAGCCAAAGCTGCCTGAGGGACCGCCGAATTGAGGGTGTG
BG8-B12    ................................................................................................
                1710      1720      1730      1740      1750      1760      1770      1780      1790      1800
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6BBGc-1    CGACCTCCAACTCAAAGCCAATTGGAAGAAAGAAACCATAGAAAGGAAGGAAAGGGGAGGGAGACAGAGATCCTGGAAGAGATATGGGCATTTGGGGAAA
BG8-B12    ................................................................................................
                1810      1820      1830      1840      1850      1860      1870      1880      1890      1900
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
6BBGc-1    TAGTGTGACCGTGTATCAGGCTTTGTGGACATCTAACGAATATGTCATGTTTTTGTAAATACAAGCATGCACGCAGAAACAAAGGCAGAAAACTGCTTTG
BG8-B12    ...........................................................................T....................

           ....|..
6BBGc-1    GGTGTTA
BG8-B12    .......
```

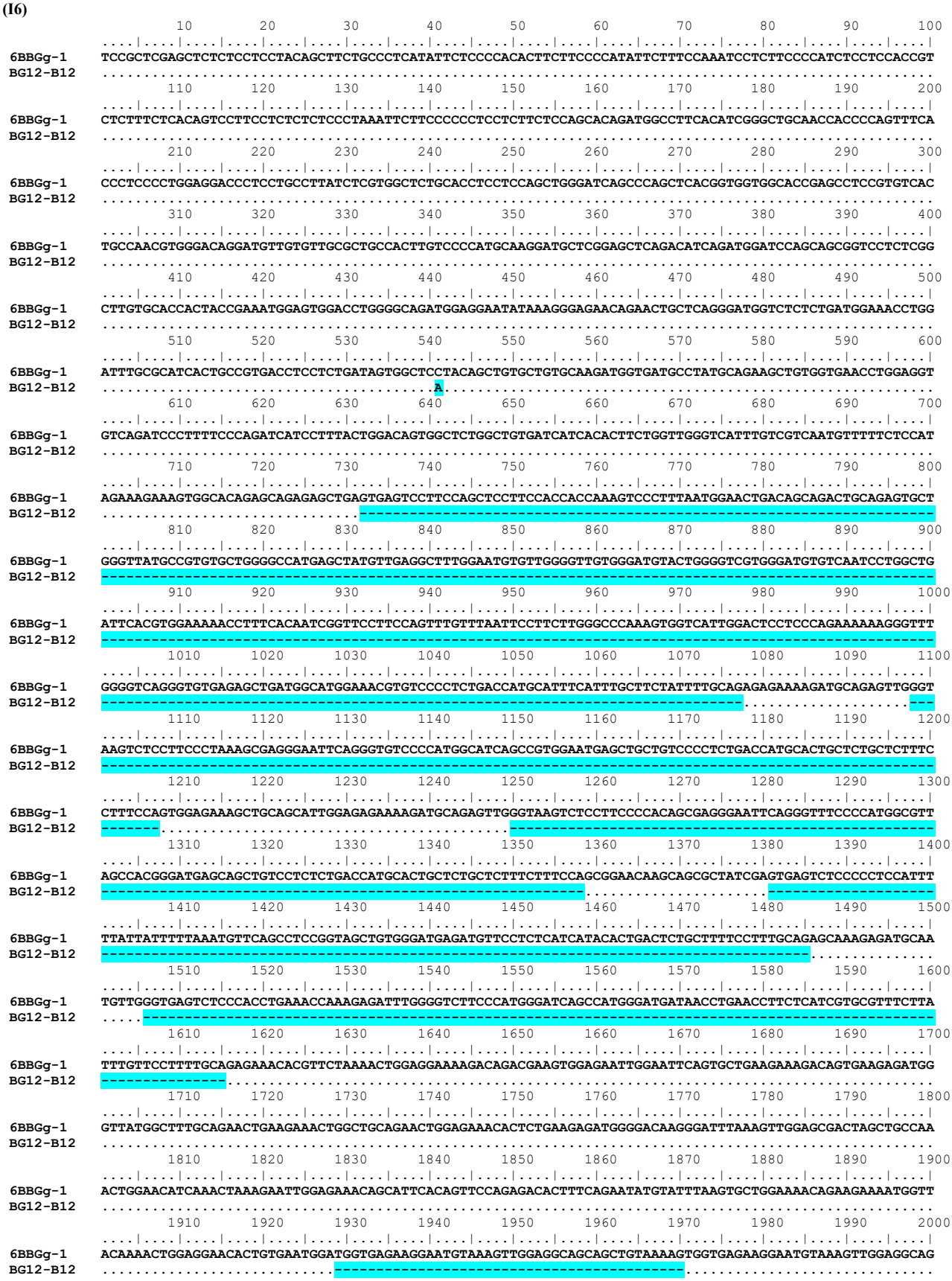
(I4)

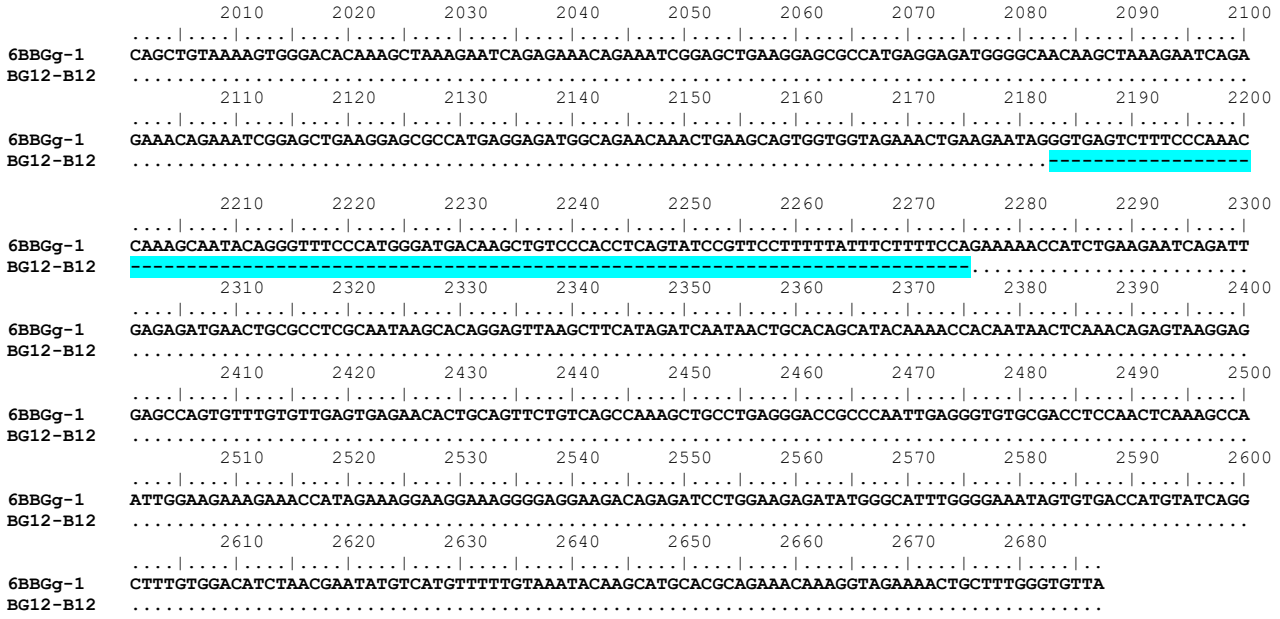


(I5)


10 20 30 40 50 60 70 80 90 100
6BBGf-1 TCCGCTCGAGCTCTCTCCTCCTACAGTTTCTGCCCTCATATTCTCCCCACACTTCTTCCCCATATTCTTTCCAAATCCTCTTCAGCATCTTCTCCACCAT
BG5-B12 .....T.....
110 120 130 140 150 160 170 180 190 200
6BBGf-1 CTCCTTCTCAGTATCCTTCCTCTCTCTCCCTAAATTCTTCTCCCCTTCTCTTCTCCAGTACAGATGGCCTTCACATCGGGCTGCAACCACCCCAGATTCA
BG5-B12 .....
210 220 230 240 250 260 270 280 290 300
6BBGf-1 CCCTCCCCTGGAGGACACTCCTGCCTTATCTCCTGGCTCTGCACCTCCTCCAGCCGGGATCAGCTCAGCTTACAGTGGTAGCACCGAGCCTCCGTGTCAC
BG5-B12 .....
310 320 330 340 350 360 370 380 390 400
6BBGf-1 TGCCATCGTGGGACAGGATGTCGTGCTGCGCTGCCACTTGTCCCCTTGCAAGGATGCTTGGAGCTCAGACATCAGATGGATACAGCTGCGGTCCTCTGGG
BG5-B12 .....
410 420 430 440 450 460 470 480 490 500
6BBGf-1 ATTGTGCACCACTACCAAAATGGAGAGGACCTGGAGCAAATGCAGGAATATAAAGGGAGAACAGAACTGCTCAGGGATGGTCTCTATGATGGAAACCTGG
BG5-B12 .....
510 520 530 540 550 560 570 580 590 600
6BBGf-1 ATTTGCGCATCACTGCTGTGAGCTCCTCCGATAGTGGCTCATACAGCTGTGCTGTGCAAGATGGTGATGGCTATGCAGACGCTGTGGTGGACCTGGAGGT
BG5-B12 .....
610 620 630 640 650 660 670 680 690 700
6BBGf-1 GTCAGATCCCTTTTCCCAGATCCTTCATCCCTGGAAGGTGGCTCTGGCTGTGATCGTCACACTTCTGGTTGGGTCATTTGTCATCATTGCTTTTCTCTAT
BG5-B12 .....
710 720 730 740 750 760 770 780 790 800
6BBGf-1 AGGAAGAAAGCGGCACAGAGCAGAGAGCTGAAGGGGAAAGATGCAGCACTGGCGGAACTACCTGCGATATTGGGTGTATGTGCTGCAAATCTGAAGATCC
BG5-B12 .....
810 820 830 840 850 860 870 880 890 900
6BBGf-1 TAGCTTCAAAACTGATGAAACAAATGGAAAAATTGGAGATTCAGAATTCAGTCTTGGAGAAATGGTATGAGAATACGGAGGAACTGGCTGCAGATCTGGA
BG5-B12 .....
910 920 930 940 950 960 970 980 990 1000
6BBGf-1 GGAACATCTTGCTGAGAAGGATTTAAGCACTGCAGATCTGAAGCTACTAGCTGCAAAACTGGTGGAACAAAGAGAAGCAGTGGAGGAATGGGATTCACAG
BG5-B12 .....
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
6BBGf-1 CTGAGGAAACAGTATGAAAAGTTGGGTTCGTGTGCTGCAAATCTGAAGACACAACTTAAAAAGTTGGAGAACGAAATTGAAGAAGTGGAGAAACACCTTA
BG5-B12 .....
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
6BBGf-1 AAAAGATTGGTATACGTGCTCCTAATCTGAGGCTACACATGGCAGAACTGGTGGATCAAGTTGAAGCAGTGGAGAATCGGAAATCAGAGTGGAAGAGATA
BG5-B12 .....
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
6BBGf-1 TTTGACAAATATAGGTTTACGTGCTGCAGAGCTGAAAAAAAACGTTGCAGAACTGAAGAAACGAATTGGAGCATTGGAAACTAAAGAATTGGAAAAACCA
BG5-B12 .....
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
6BBGf-1 TCTAAAGAACAGGATTGAGAGATGAACTGCGCCTCACAGTAACCACAGGAGATAAGCTTCATAGACTGCTGATTGCACAGGATAGCAACATCGCCATAAC
BG5-B12 .....
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
6BBGf-1 GCAAAGCAAGAAAGGAAATCCACACGGGGAACAAGAGGAGCCAGTGTTTGTATTGAGTGAGAACACTGCAGTTCTGCAAGCCACAACTGCCTGAGGGACC
BG5-B12 .....
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
6BBGf-1 ACCAAACTGAGGGTGTGTGACCTCCATCTCAAATCCAGTTGGAAGAAAAAAACCATAGAAAAGAAGACTACAAGAGGAAGACAGAGATCCTGGGAAAGGG
BG5-B12 .....G.....
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
6BBGf-1 ACAGACATTTTGGGAATTAACATGGCCATGTATCAGGGGTGAGGAATTCTAATGAATATGTAAGGCTTCTGGAAATATAAACATGCACACAGAAGTAAAG
BG5-B12 .....
1710 1720
6BBGf-1 GTAGAAAACTGCTTTGGGTGTTA
BG5-B12 .....

(I6)

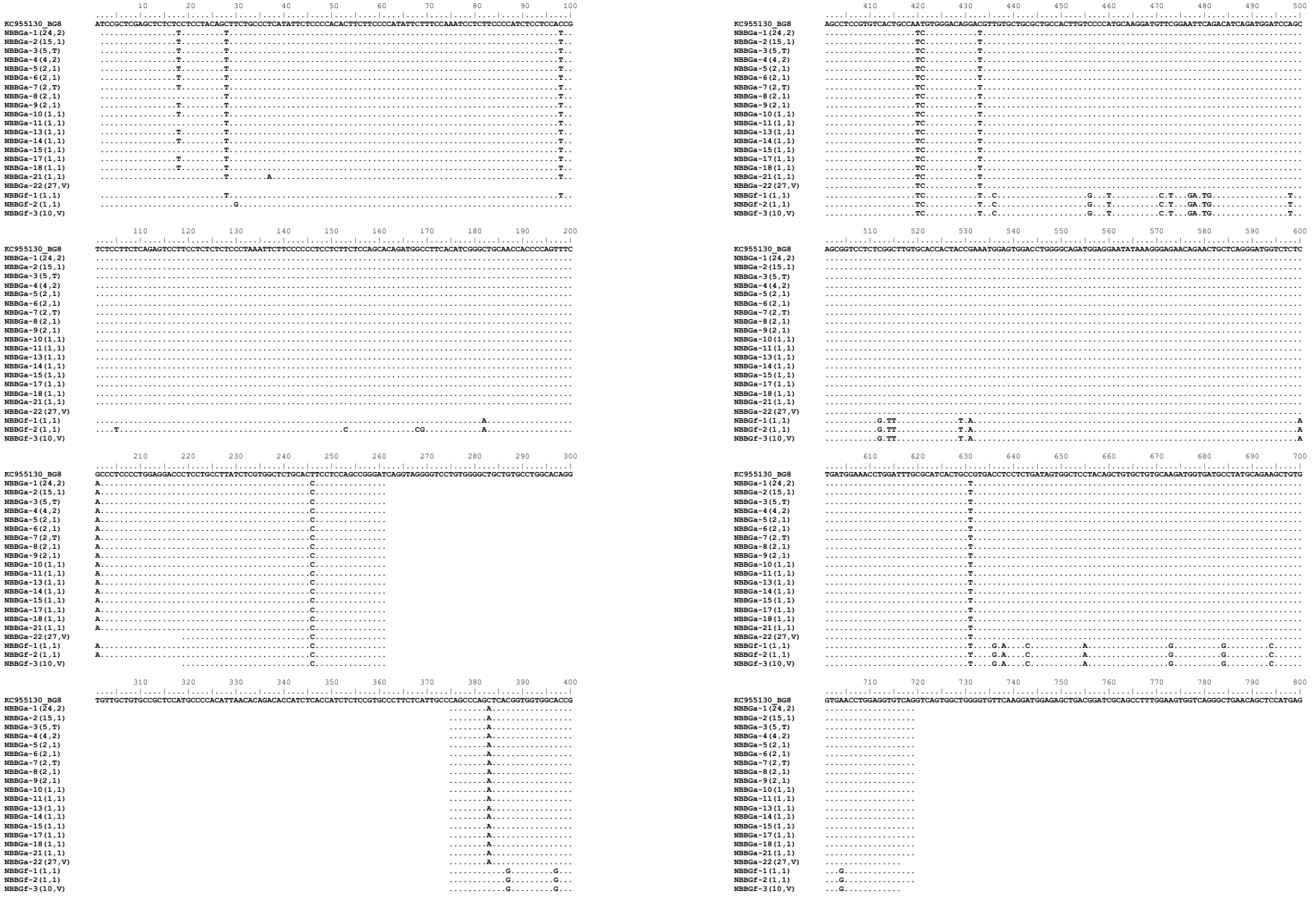


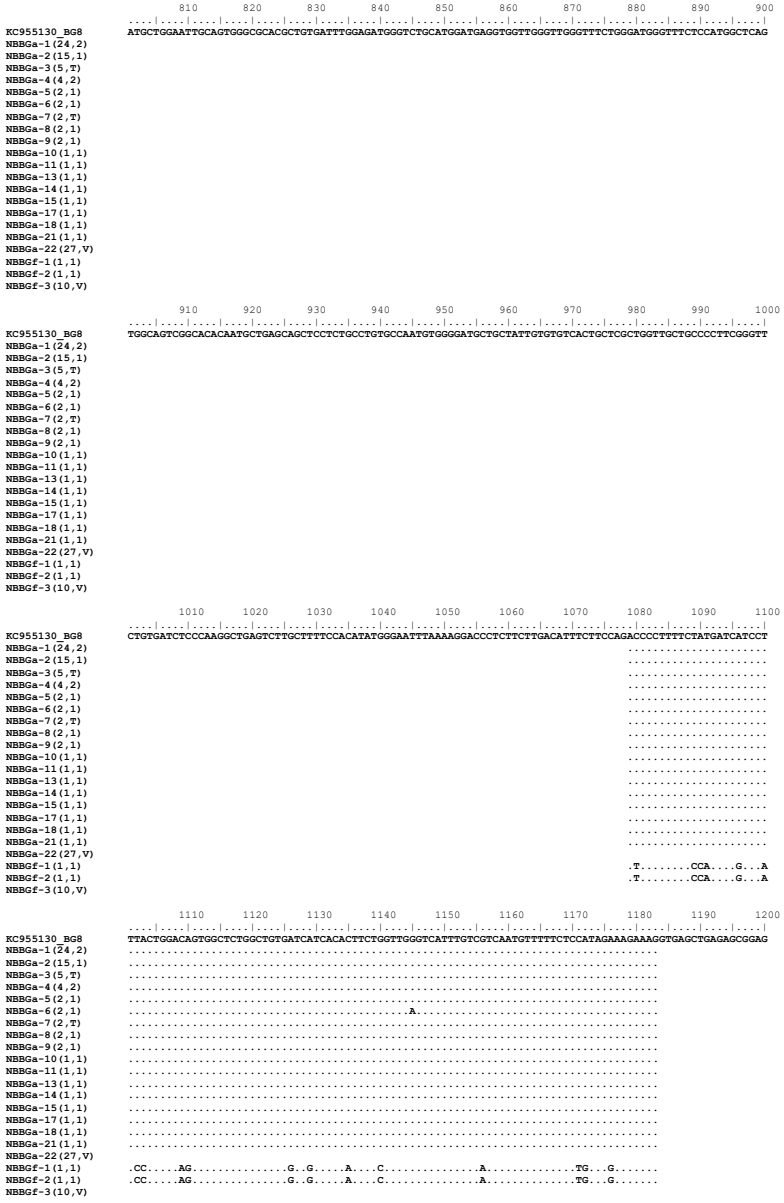


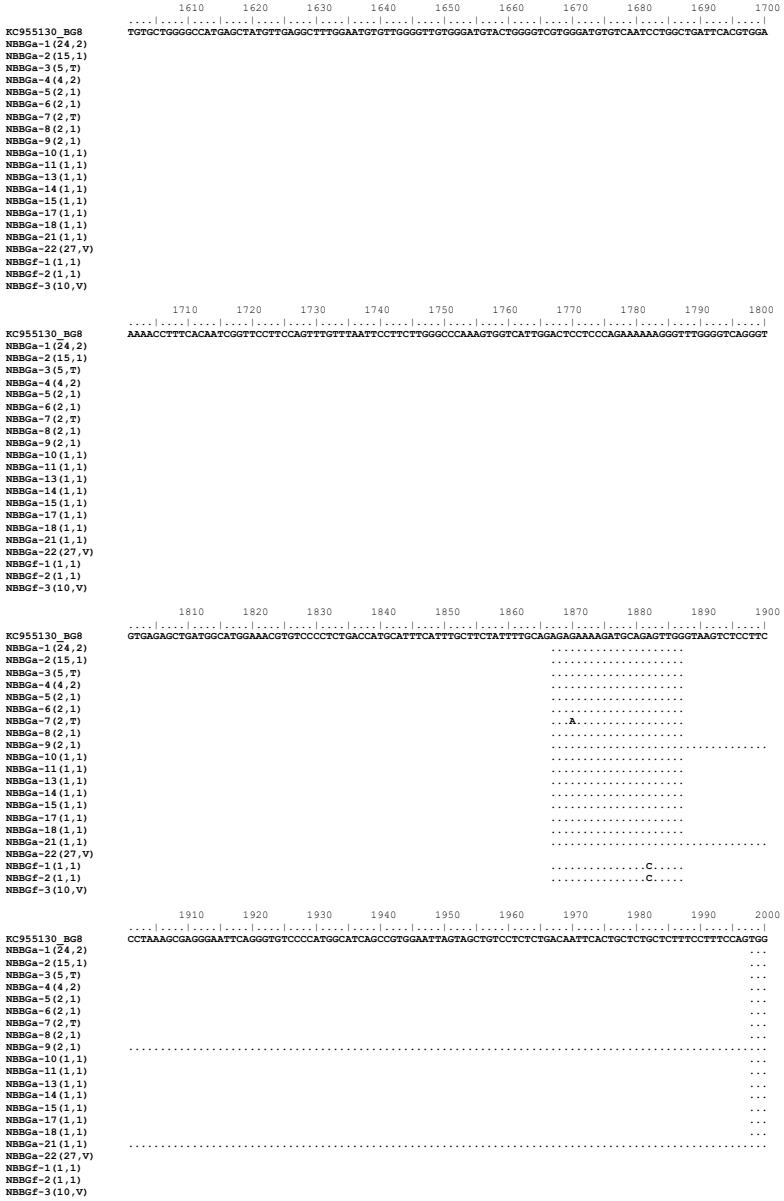
Appendix I. Nucleotide sequence alignments show that six BG genes found from B cells of line 6_1 (B2) are the same genes as the ones in B12 haplotype. Names of the transcripts follow the convention: abbreviated line name, "B" for B cells, "BG", a letter representing the exon 2 sequence, a dash and then a number representing the alternative splicing variant with "1" being the most frequently detected clone. Names of the gene follow the convention: "BG", the number of the gene locus for the B12 haplotype, a dash and the chicken haplotype B12. Letters indicate nucleotides, dots indicate identities with the top sequences, dashes indicate no sequence present compared to the top sequences. The differences between the major cDNA sequences from the six BG genes of B2 and their counterparts from B12 are labeled with blue colour. The differences are either due to one nucleotide PCR artifact in (I5) and (I6) or the intron retention in (I1), (I3), (I4) and (I6). It should be noted that the B12 BG cDNA sequences are annotated from genomic sequences, which may or may not represent the expressed transcripts.

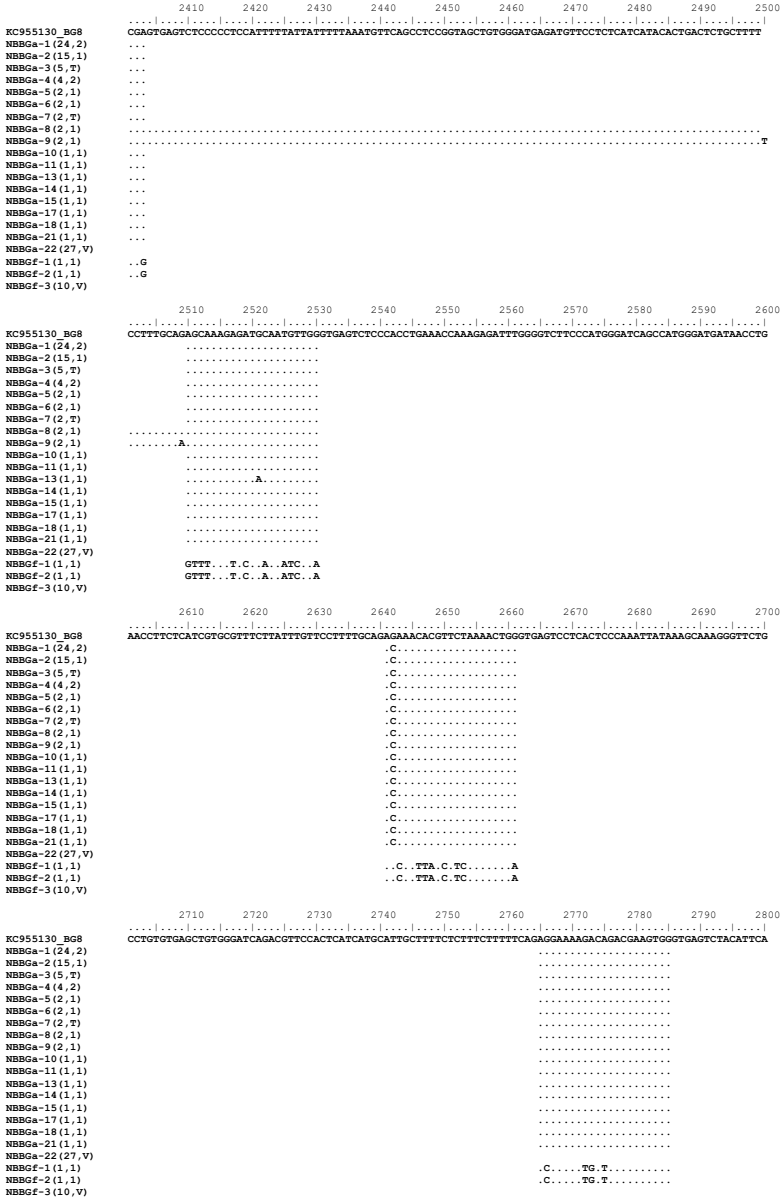
Appendix J.

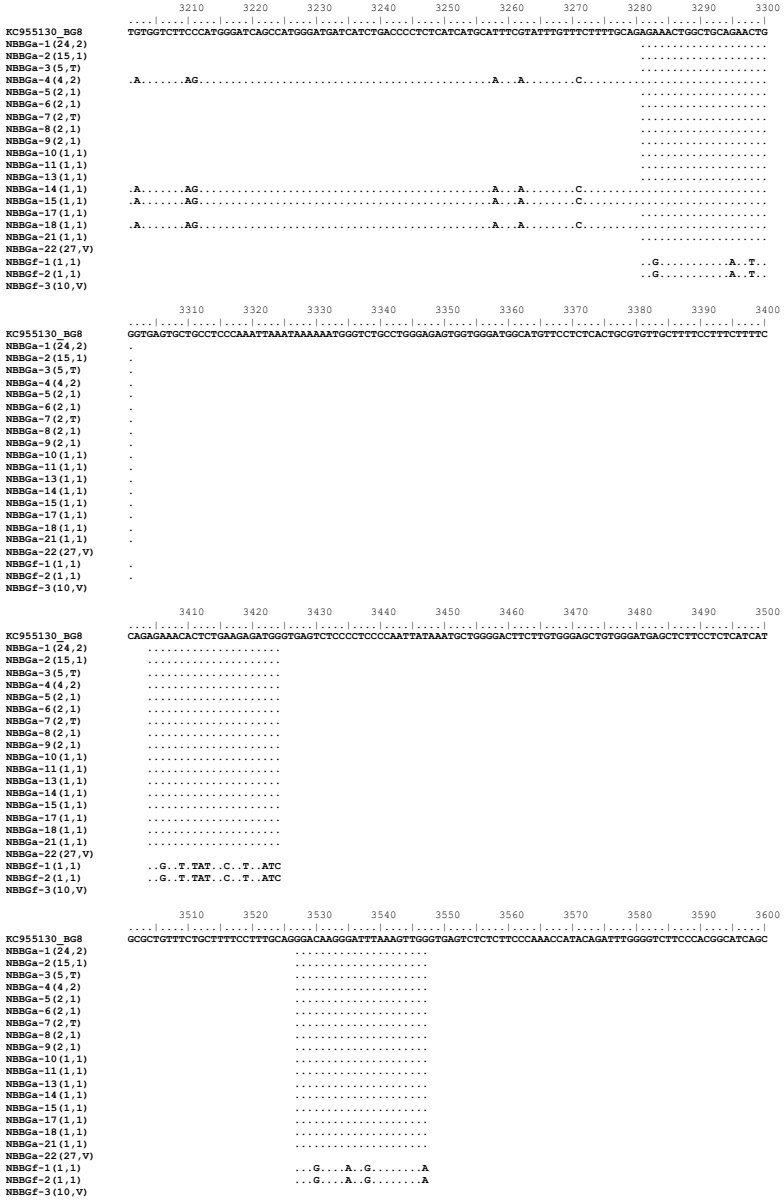
(J1)

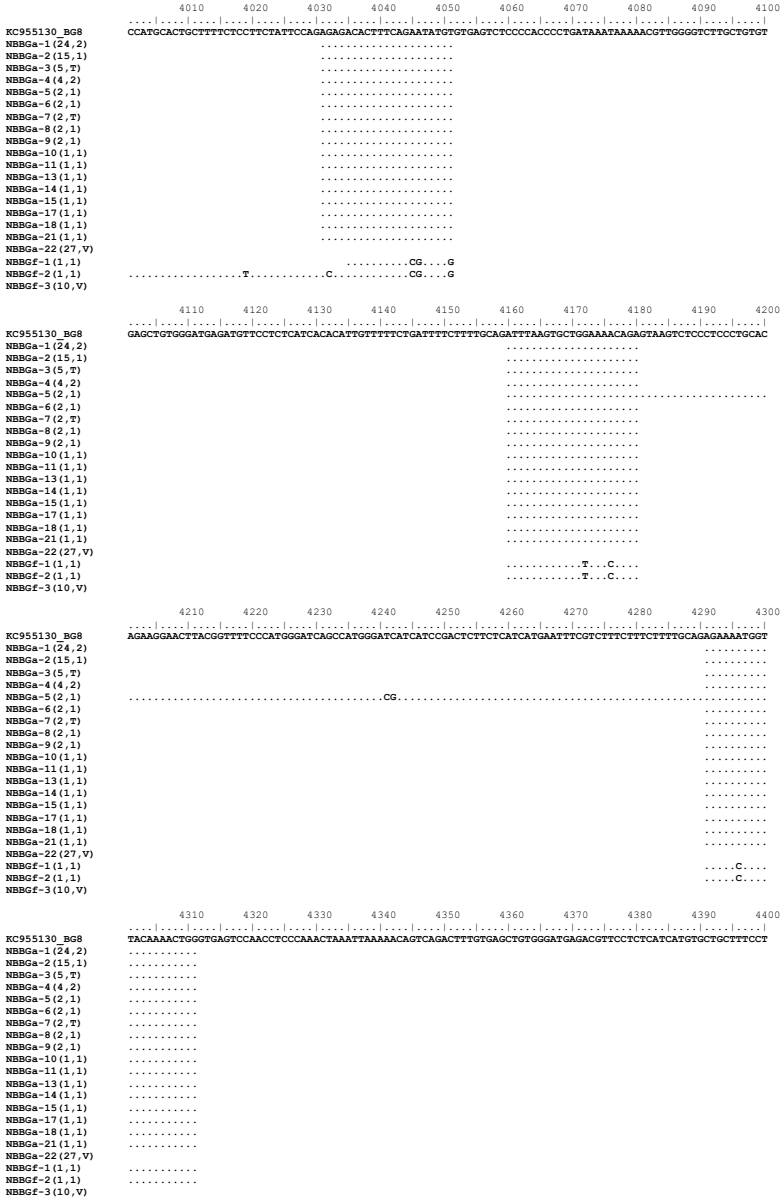


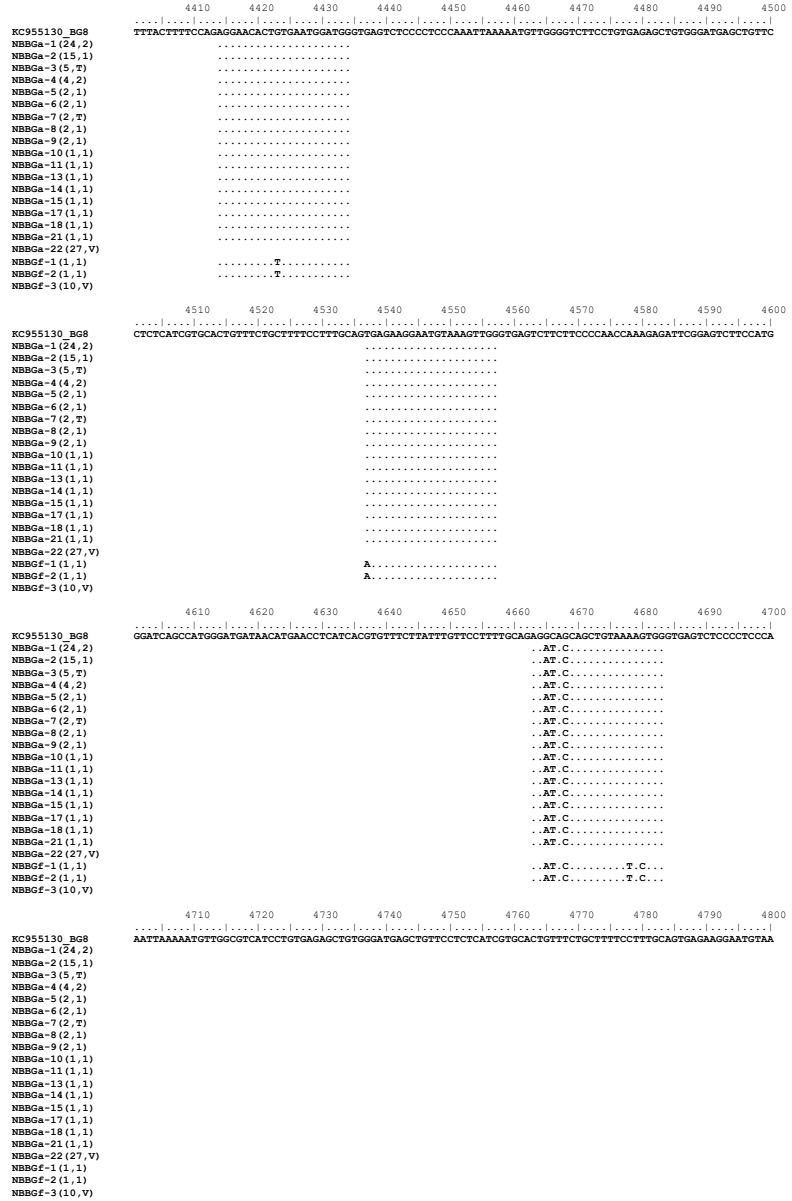

810 820 830 840 850 860 870 880 890 900
KC955130_BG8 ATGCTGGAATTGCAGTGGGCGCACGCTGTGATTTGGAGATGGGTCTGCATGGATGAGGTGGTTGGGTTTCTGGGATGGGTTTCTCCATGGCTCAG
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
910 920 930 940 950 960 970 980 990 1000
KC955130_BG8 TGGCAGTCGGCACACAATGCTGAGCAGCTCCTCTGCCTGTGCCAATGTGGGGATGCTGCTATTGTGTGTCACTGCTCGCTGGTTGCTGCCCCTTCGGGTT
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
KC955130_BG8 CTGTGATCTCCCAAGGCTGAGTCTTGCTTTTCCACATATGGGAATTTAAAAGGACCCTCTTCTTGACATTTCTTCCAGACCCCTTTTCTATGATCATCCT
NBBGa-1(24,2) ....
NBBGa-2(15,1) ....
NBBGa-3(5,T) ....
NBBGa-4(4,2) ....
NBBGa-5(2,1) ....
NBBGa-6(2,1) ....
NBBGa-7(2,T) ....
NBBGa-8(2,1) ....
NBBGa-9(2,1) ....
NBBGa-10(1,1) ....
NBBGa-11(1,1) ....
NBBGa-13(1,1) ....
NBBGa-14(1,1) ....
NBBGa-15(1,1) ....
NBBGa-17(1,1) ....
NBBGa-18(1,1) ....
NBBGa-21(1,1) ....
NBBGa-22(27,V)
NBBGf-1(1,1) .T........CCA....G...A
NBBGf-2(1,1) .T........CCA....G...A
NBBGf-3(10,V)
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
KC955130_BG8 TTACTGGACAGTGGCTCTGGCTGTGATCATCACACTTCTGGTTGGGTCATTTGTCGTCAATGTTTTTCTCCATAGAAAGAAAGGTGAGCTGAGAGCGGAG
NBBGa-1(24,2) ....
NBBGa-2(15,1) ....
NBBGa-3(5,T) ....
NBBGa-4(4,2) ....
NBBGa-5(2,1) ....
NBBGa-6(2,1) ....A.....
NBBGa-7(2,T) ....
NBBGa-8(2,1) ....
NBBGa-9(2,1) ....
NBBGa-10(1,1) ....
NBBGa-11(1,1) ....
NBBGa-13(1,1) ....
NBBGa-14(1,1) ....
NBBGa-15(1,1) ....
NBBGa-17(1,1) ....
NBBGa-18(1,1) ....
NBBGa-21(1,1) ....
NBBGa-22(27,V)
NBBGf-1(1,1) .CC.....AG.............G..G.....A....C.............A.............TG...G.......
NBBGf-2(1,1) .CC.....AG.............G..G.....A....C.............A.............TG...G.......
NBBGf-3(10,V)
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
KC955130_BG8 GGGATGGAGCACAGGGAGGTGTTGTGCATGGACAGGGATGGTCGGGGTGGTGCTGAGCTGTGGTCCACGGAGGTACACAGGTGGAGGAACCGTGACTTTT
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
KC955130_BG8 CATGGGATTCCCAGTGCTCATTAAATAACATTTGCCTTTCTTTTGGGGAATAAAAGAAGGGGAAAAACGATAGTGGTAAGGGTGGGCAGATAGGAATGTG
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
KC955130_BG8 GCTGGACTGTGGGGCAGGTGGAAAGTCCAAACCCTCTGGAGAAGTCCCCACAAACCAAGCTGCCCTGCTGACCAGCTATTTCTCTGCTTTGTTTTCCAGT
NBBGa-1(24,2) .
NBBGa-2(15,1) .
NBBGa-3(5,T) .
NBBGa-4(4,2) .
NBBGa-5(2,1) .
NBBGa-6(2,1) .
NBBGa-7(2,T) .
NBBGa-8(2,1) .
NBBGa-9(2,1) .
NBBGa-10(1,1) .
NBBGa-11(1,1) .
NBBGa-13(1,1) .
NBBGa-14(1,1) .
NBBGa-15(1,1) .
NBBGa-17(1,1) .
NBBGa-18(1,1) .
NBBGa-21(1,1) .
NBBGa-22(27,V)
NBBGf-1(1,1) C
NBBGf-2(1,1) C
NBBGf-3(10,V)
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
KC955130_BG8 GGCACAGAGCAGAGAGCTGAGTGAGTCCTTCCATCCCCATCCACCACCAAAGTCCCTTTAATGGAACTGACAGCAGACTGCAGAGTGCTGGGTTATGCCA
NBBGa-1(24,2) ....
NBBGa-2(15,1) ....
NBBGa-3(5,T) ....
NBBGa-4(4,2) ....
NBBGa-5(2,1) ....
NBBGa-6(2,1) ....
NBBGa-7(2,T) ....
NBBGa-8(2,1) ....
NBBGa-9(2,1) ....
NBBGa-10(1,1) ....A........
NBBGa-11(1,1) ....
NBBGa-13(1,1) ....
NBBGa-14(1,1) ....
NBBGa-15(1,1) ....
NBBGa-17(1,1) ....
NBBGa-18(1,1) ....
NBBGa-21(1,1) ....
NBBGa-22(27,V)
NBBGf-1(1,1) ..............................G...T..............................T..A.................
NBBGf-2(1,1) ....
NBBGf-3(10,V)

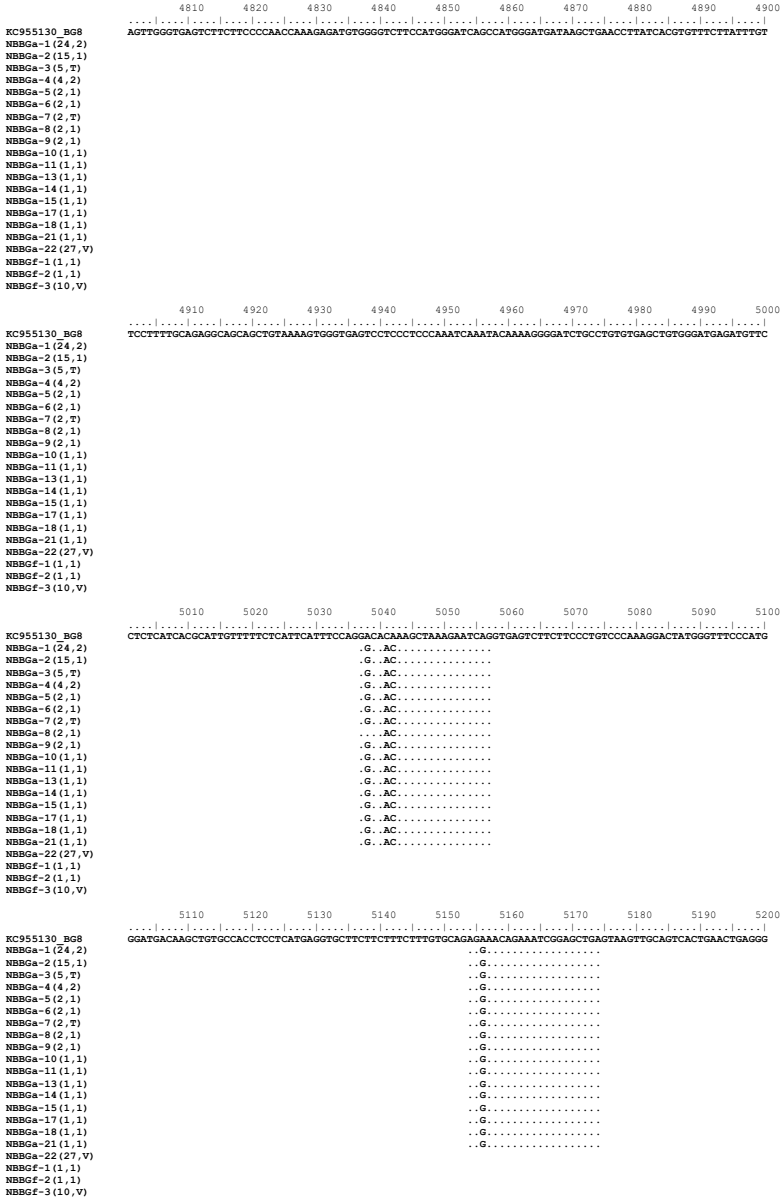

1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
KC955130_BG8 TGTGCTGGGGCCATGAGCTATGTTGAAGCTTTGGAATGTGTTGGGGTTGTGGGATGTACTGGGGTCGTGGGATGTGTCAATCCTGGCTGATTCACGTGGA
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
KC955130_BG8 AAAACCTTTCACAATCGGTTCCTTCCAGTTTGTTTAATTCCTTCTTGGGCCCAAAGTGGTCATTGGACTCCTCCCAGAAAAAAGGGTTTGGGGTCAGGGT
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
KC955130_BG8 GTGAGAGCTGATGGCATGGAAACGTGTCCCCTCTGACCATGCATTTCATTTGCTTCTATTTTGCAGAGAGAAAAGATGCAGAGTTGGGTAAGTCTCCTTC
NBBGa-1(24,2) ....
NBBGa-2(15,1) ....
NBBGa-3(5,T) ....
NBBGa-4(4,2) ....
NBBGa-5(2,1) ....
NBBGa-6(2,1) ....
NBBGa-7(2,T) ...A......
NBBGa-8(2,1) ....
NBBGa-9(2,1) ....
NBBGa-10(1,1) ....
NBBGa-11(1,1) ....
NBBGa-13(1,1) ....
NBBGa-14(1,1) ....
NBBGa-15(1,1) ....
NBBGa-17(1,1) ....
NBBGa-18(1,1) ....
NBBGa-21(1,1) ....
NBBGa-22(27,V)
NBBGf-1(1,1) ....C.....
NBBGf-2(1,1) ....C.....
NBBGf-3(10,V)
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
KC955130_BG8 CCTAAAGCGAGGGAATTCAGGGTGTCCCCATGGCATCAGCCGTGGAATTAGTAGCTGTCCTCTCTGACAATTCACTGCTCTGCTCTTTCCTTTCCAGTGG
NBBGa-1(24,2) ...
NBBGa-2(15,1) ...
NBBGa-3(5,T) ...
NBBGa-4(4,2) ...
NBBGa-5(2,1) ...
NBBGa-6(2,1) ...
NBBGa-7(2,T) ...
NBBGa-8(2,1) ...
NBBGa-9(2,1) ....
NBBGa-10(1,1) ...
NBBGa-11(1,1) ...
NBBGa-13(1,1) ...
NBBGa-14(1,1) ...
NBBGa-15(1,1) ...
NBBGa-17(1,1) ...
NBBGa-18(1,1) ...
NBBGa-21(1,1) ....
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
KC955130_BG8 AGAAAGCTGCAGCATTGGGTGAGTTATATTCCCCAAGCCAAAGTACTTTGGGTCTTCCCATTGGAAGTTATTTCCTCAGACCATCCTTTCTGTTGTGTTT
NBBGa-1(24,2) ....
NBBGa-2(15,1) ....
NBBGa-3(5,T) ....
NBBGa-4(4,2) ....
NBBGa-5(2,1) ....
NBBGa-6(2,1) ....
NBBGa-7(2,T) ....
NBBGa-8(2,1) ....
NBBGa-9(2,1) ....
NBBGa-10(1,1) ....
NBBGa-11(1,1) ....
NBBGa-13(1,1) ....
NBBGa-14(1,1) ....
NBBGa-15(1,1) ....
NBBGa-17(1,1) ....
NBBGa-18(1,1) ....
NBBGa-21(1,1) ....
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
KC955130_BG8 GCTTTGGCATCATGTTAGTAAAATGCCTTCTTGGGACCAAAGTGGTCATTGGCCACTTCCCAGAAAAAAAGATTTGGGGGCAGGGTGTGGGAGCTGATGG
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1) ..........G.......T..........
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1) ..........G.......T..........A
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
KC955130_BG8 CATGGAAATTTGTCCCCTCTGACCATGCTTTTCCTTTGCTTCTTTTTGCAGAGAGAAAAGATGCAGAGTTGGGTAAGTCTCCTTCCCCACAGTGAGGGAA
NBBGa-1(24,2) ....
NBBGa-2(15,1) ....
NBBGa-3(5,T) ....
NBBGa-4(4,2) ....
NBBGa-5(2,1) ....
NBBGa-6(2,1) ....
NBBGa-7(2,T) ....
NBBGa-8(2,1) ....
NBBGa-9(2,1) .....CA...T.....A.....C.....
NBBGa-10(1,1) ....
NBBGa-11(1,1) ....
NBBGa-13(1,1) ....
NBBGa-14(1,1) ....
NBBGa-15(1,1) ....
NBBGa-17(1,1) ....
NBBGa-18(1,1) ....
NBBGa-21(1,1) .....CA...T.....A.....C.....
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
KC955130_BG8 TTCAGGGTTTCCCCATGGCGTTAGCCACGGGATGGGCAGCTGTCCTCTCTGACCATGCACTGCTCTGCTCTTTCTTTTCCAGCGGAACAAGCAGCGCTAT
NBBGa-1(24,2) ....A..
NBBGa-2(15,1) ....A..
NBBGa-3(5,T) ....A..
NBBGa-4(4,2) ....A..
NBBGa-5(2,1) ....A..
NBBGa-6(2,1) ....A..
NBBGa-7(2,T) ....A..
NBBGa-8(2,1) ....A..
NBBGa-9(2,1) ....A..
NBBGa-10(1,1) ....A..
NBBGa-11(1,1) ....A..
NBBGa-13(1,1) ....A..
NBBGa-14(1,1) ....A..
NBBGa-15(1,1) ....A..
NBBGa-17(1,1) ....A..
NBBGa-18(1,1) ....A..
NBBGa-21(1,1) ....A..
NBBGa-22(27,V)
NBBGf-1(1,1) ......T..AT.A.A...
NBBGf-2(1,1) ......T..AT.A.A...
NBBGf-3(10,V)

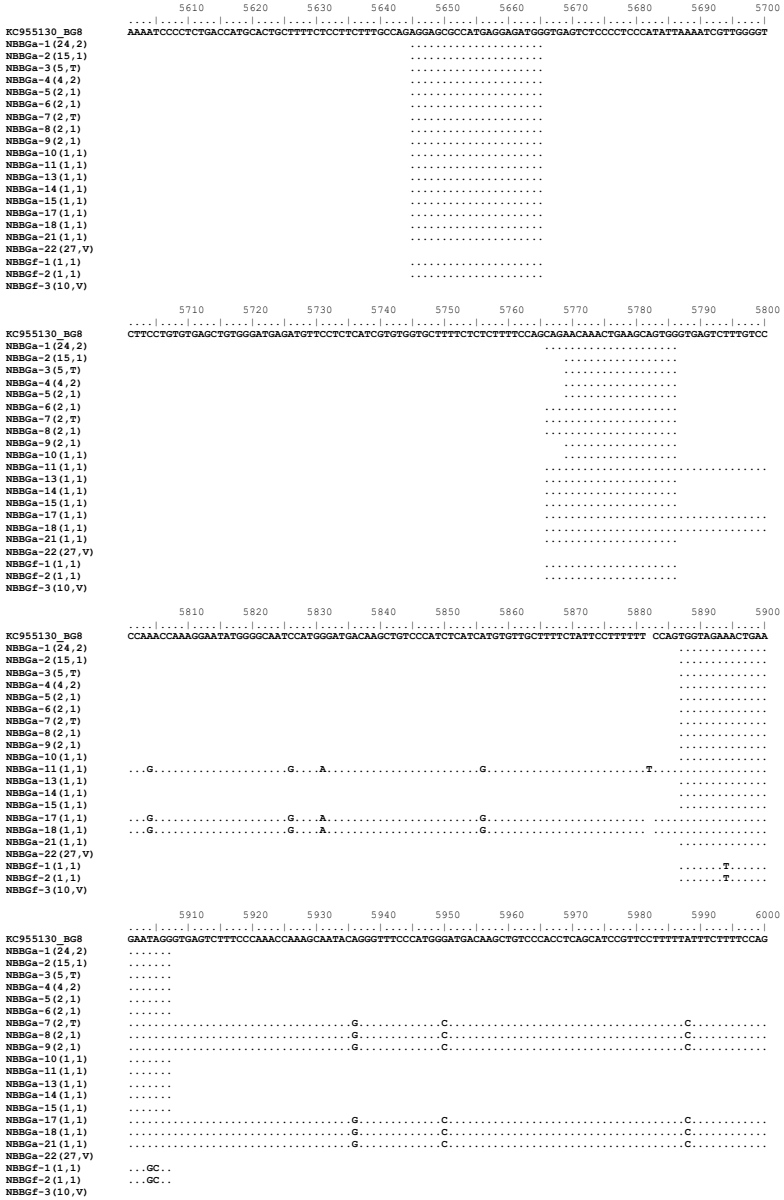

2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
KC955130_BG8 CGAGTGAGTCTCCCCCTCCATTTTTATTATTTTTAAATGTTCAGCCTCCGGTAGCTGTGGGATGAGATGTTCCTCTCATCATACACTGACTCTGCTTTT
NBBGa-1(24,2) ...
NBBGa-2(15,1) ...
NBBGa-3(5,T) ...
NBBGa-4(4,2) ...
NBBGa-5(2,1) ...
NBBGa-6(2,1) ...
NBBGa-7(2,T) ...
NBBGa-8(2,1) .....
NBBGa-9(2,1) .....T
NBBGa-10(1,1) ...
NBBGa-11(1,1) ...
NBBGa-13(1,1) ...
NBBGa-14(1,1) ...
NBBGa-15(1,1) ...
NBBGa-17(1,1) ...
NBBGa-18(1,1) ...
NBBGa-21(1,1) ...
NBBGa-22(27,V)
NBBGf-1(1,1) ..G
NBBGf-2(1,1) ..G
NBBGf-3(10,V)
2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
KC955130_BG8 CCTTTGCAGAGCAAAGAGATGCAATGTTGGGTGAGTCTCCCACCTGAAACCAAAGAGATTTGGGGTCTTCCCATGGGATCAGCCATGGGATGATAACCTG
NBBGa-1(24,2) .....
NBBGa-2(15,1) .....
NBBGa-3(5,T) .....
NBBGa-4(4,2) .....
NBBGa-5(2,1) .....
NBBGa-6(2,1) .....
NBBGa-7(2,T) .....
NBBGa-8(2,1) .....
NBBGa-9(2,1) .....A.....
NBBGa-10(1,1) .....
NBBGa-11(1,1) .....
NBBGa-13(1,1) .....A.....
NBBGa-14(1,1) .....
NBBGa-15(1,1) .....
NBBGa-17(1,1) .....
NBBGa-18(1,1) .....
NBBGa-21(1,1) .....
NBBGa-22(27,V)
NBBGf-1(1,1) GTTT...T.C..A..ATC..A
NBBGf-2(1,1) GTTT...T.C..A..ATC..A
NBBGf-3(10,V)
2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
KC955130_BG8 AACCTTCTCATCGTGGGTTTCTTATTTGTTCCTTTTGCAGAGAAACACGTTCTAAAACTGGGTGAGTCCTCACTCCCAAATTATAAAGCAAAGGGTTCTG
NBBGa-1(24,2) .C.....
NBBGa-2(15,1) .C.....
NBBGa-3(5,T) .C.....
NBBGa-4(4,2) .C.....
NBBGa-5(2,1) .C.....
NBBGa-6(2,1) .C.....
NBBGa-7(2,T) .C.....
NBBGa-8(2,1) .C.....
NBBGa-9(2,1) .C.....
NBBGa-10(1,1) .C.....
NBBGa-11(1,1) .C.....
NBBGa-13(1,1) .C.....
NBBGa-14(1,1) .C.....
NBBGa-15(1,1) .C.....
NBBGa-17(1,1) .C.....
NBBGa-18(1,1) .C.....
NBBGa-21(1,1) .C.....
NBBGa-22(27,V)
NBBGf-1(1,1) ..C..TTA.C.TC.....A
NBBGf-2(1,1) ..C..TTA.C.TC.....A
NBBGf-3(10,V)
2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
KC955130_BG8 CCTGTGTGAGCTGTGGGATCAGACGTTCCACTCATCATGCATTGCTTTTCTCTTTCTTTTTCAGAGGAAAAGACAGACGAAGTGGGTGAGTCTACATTCA
NBBGa-1(24,2) .....
NBBGa-2(15,1) .....
NBBGa-3(5,T) .....
NBBGa-4(4,2) .....
NBBGa-5(2,1) .....
NBBGa-6(2,1) .....
NBBGa-7(2,T) .....
NBBGa-8(2,1) .....
NBBGa-9(2,1) .....
NBBGa-10(1,1) .....
NBBGa-11(1,1) .....
NBBGa-13(1,1) .....
NBBGa-14(1,1) .....
NBBGa-15(1,1) .....
NBBGa-17(1,1) .....
NBBGa-18(1,1) .....
NBBGa-21(1,1) .....
NBBGa-22(27,V)
NBBGf-1(1,1) .C.....TG.T.....
NBBGf-2(1,1) .C.....TG.T.....
NBBGf-3(10,V)
2810 2820 2830 2840 2850 2860 2870 2880 2890 2900
KC955130_BG8 CTAAAGCAAAGAAATATGGGGTCTCCCATGGGATGACAAGCTGTCCCAAAAAATCATGTGGTGCTTTTTCTTGTCTTTTTATTTATTTATTTATTTATTT
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
KC955130_BG8 ATTTGCAGAGAATTGGAATTCAGTGCTGAGTAAGTTGCAGTCACTGAACTGAGGGAATGTGGGGTCTTCCTAAGGGACTGCGTAGGGGAGAAGTTCCCAT
NBBGa-1(24,2) .....
NBBGa-2(15,1) .....
NBBGa-3(5,T) .....
NBBGa-4(4,2) .....
NBBGa-5(2,1) .....
NBBGa-6(2,1) .....
NBBGa-7(2,T) .....
NBBGa-8(2,1) .....
NBBGa-9(2,1) .....
NBBGa-10(1,1) .....
NBBGa-11(1,1) .....
NBBGa-13(1,1) .....
NBBGa-14(1,1) .....
NBBGa-15(1,1) .....
NBBGa-17(1,1) .....
NBBGa-18(1,1) .....
NBBGa-21(1,1) .....
NBBGa-22(27,V)
NBBGf-1(1,1) ..G...C.......A.....
NBBGf-2(1,1) ..G...C.......A.....
NBBGf-3(10,V)
3010 3020 3030 3040 3050 3060 3070 3080 3090 3100
KC955130_BG8 GCACTGCTTTTCTCTTTCTTTTCCAGAGAAAAGACAGTGAAGAGATGGGTGAGTCTCTCCTCCCAAATTAAAAACGTTGGGGTTCCCATGTGGGAGCTGTG
NBBGa-1(24,2) .A.....
NBBGa-2(15,1) .A.....
NBBGa-3(5,T) .A.....
NBBGa-4(4,2) .A.....
NBBGa-5(2,1) .A.....
NBBGa-6(2,1) .A.....
NBBGa-7(2,T) .A.....
NBBGa-8(2,1) .A.....
NBBGa-9(2,1) .A.....
NBBGa-10(1,1) .A.....
NBBGa-11(1,1) .A.....
NBBGa-13(1,1) .A.....
NBBGa-14(1,1) .A.....
NBBGa-15(1,1) .A.....
NBBGa-17(1,1) .A.....
NBBGa-18(1,1) .A.....
NBBGa-21(1,1) .A.....
NBBGa-22(27,V)
NBBGf-1(1,1) ....T.....
NBBGf-2(1,1) ....T.....
NBBGf-3(10,V)
3110 3120 3130 3140 3150 3160 3170 3180 3190 3200
KC955130_BG8 GGATGAGATGTTCCTCTCATCAACCATCTTTTTTACTTTTCCTTTGCAGGTTATGGCTTTGCAGAACTGAGTAAGTCTCCCTCCCCAACACGGAAGGGATT
NBBGa-1(24,2) ...........G...T....
NBBGa-2(15,1) ...........G...T....
NBBGa-3(5,T) ...........G...T....
NBBGa-4(4,2) ...........G...T.....T......A...
NBBGa-5(2,1) ...........G...T....
NBBGa-6(2,1) ...........G...T....
NBBGa-7(2,T) ...........G...T....
NBBGa-8(2,1) ...........G...T....
NBBGa-9(2,1) ...........G...T....
NBBGa-10(1,1) ...........G...T....
NBBGa-11(1,1) ...........G...T....
NBBGa-13(1,1) ...........G...T....
NBBGa-14(1,1) ...........G...T.....T......A...
NBBGa-15(1,1) ...........G...T.....T......A...
NBBGa-17(1,1) ...........G...T....
NBBGa-18(1,1) ...........G...T.....T......A...
NBBGa-21(1,1) ...........G...T....
NBBGa-22(27,V)
NBBGf-1(1,1) ...C....G.......T....
NBBGf-2(1,1) ...C....G.......T....
NBBGf-3(10,V)

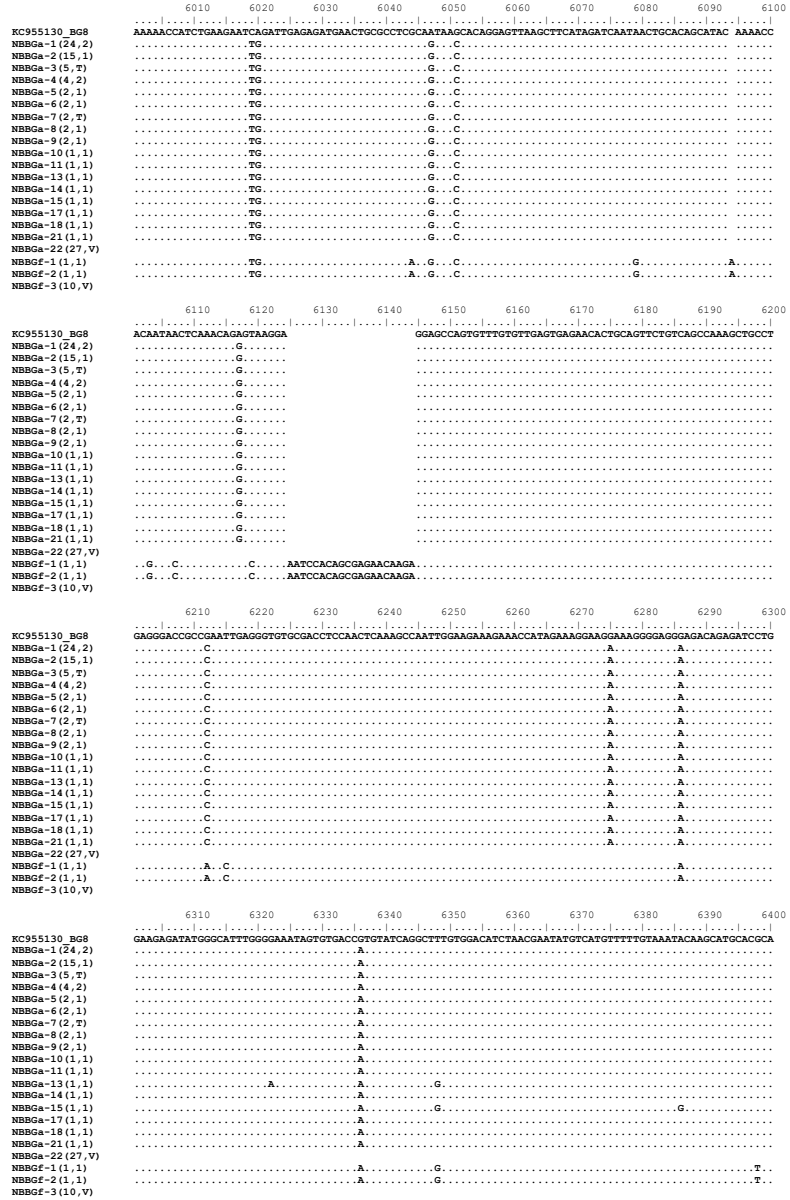

3210 3220 3230 3240 3250 3260 3270 3280 3290 3300
KC955130_BG8 TGTGGTCTTCCCATGGGATCAGCCATGGGATGATCATCTGACCCCTCTCATCATGCATTTCGTATTTGTTTCTTTTGCAGAGAAACTGGCTGCAGAACTG
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2) .A.....AG.....A...A.....C.....
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1) .A.....AG.....A...A.....C.....
NBBGa-15(1,1) .A.....AG.....A...A.....C.....
NBBGa-17(1,1)
NBBGa-18(1,1) .A.....AG.....A...A.....C.....
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1) ..G.....A..T..
NBBGf-2(1,1) ..G.....A..T..
NBBGf-3(10,V)
3310 3320 3330 3340 3350 3360 3370 3380 3390 3400
KC955130_BG8 GGTGAGTGCTGCCTCCCAAATTAAATAAAAAATGGGTCTGCCTGGGAGAGTGGTGGGATGGCATGTTCCTCTCACTGCGTGTTGCTTTTCCTTTCTTTTC
NBBGa-1(24,2) .
NBBGa-2(15,1) .
NBBGa-3(5,T) .
NBBGa-4(4,2) .
NBBGa-5(2,1) .
NBBGa-6(2,1) .
NBBGa-7(2,T) .
NBBGa-8(2,1) .
NBBGa-9(2,1) .
NBBGa-10(1,1) .
NBBGa-11(1,1) .
NBBGa-13(1,1) .
NBBGa-14(1,1) .
NBBGa-15(1,1) .
NBBGa-17(1,1) .
NBBGa-18(1,1) .
NBBGa-21(1,1) .
NBBGa-22(27,V)
NBBGf-1(1,1) .
NBBGf-2(1,1) .
NBBGf-3(10,V)
3410 3420 3430 3440 3450 3460 3470 3480 3490 3500
KC955130_BG8 CAGAGAAACACTCTGAAGAGATGGGTGAGTCTCCCCCTCCCCAATTATAAATGCTGGGGACTTCTTGTGGGAGCTGTGGGATGAGCTCTTCCTCTCATCAT
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1) ..G..T.TAT..C..T..ATC
NBBGf-2(1,1) ..G..T.TAT..C..T..ATC
NBBGf-3(10,V)
3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
KC955130_BG8 GCGCTGTTTCTGCTTTTCCTTTGCAGGGACAAGGGATTTAAAGTTGGGTGAGTCTCTCTTCCCAAACCATACAGATTTGGGGTCTTCCCACGGCATCAGC
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1) ...G....A..G.....A
NBBGf-2(1,1) ...G....A..G.....A
NBBGf-3(10,V)
3610 3620 3630 3640 3650 3660 3670 3680 3690 3700
KC955130_BG8 CATGGGATGATAATCGGACCCTTCTCATCATGCATTTCTTATTGGTTCCTTTTGCAGAGCGACTAGCTGCCAAACTGGGTGAGTCCCCCCTCCCAAATTA
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1) .TAATA.....
NBBGf-2(1,1) .TAATA.....
NBBGf-3(10,V)
3710 3720 3730 3740 3750 3760 3770 3780 3790 3800
KC955130_BG8 AATAAAAAATGGGGTCTGCCTGTGTGAGCTGTGAGATGAGATGTTCCTCTCATCATGCGCTGCTTTTCTCTTCCTTTTCCAGAACATCAAACTAAAGAAT
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1) C...A.....
NBBGf-2(1,1) C...A.....
NBBGf-3(10,V)
3810 3820 3830 3840 3850 3860 3870 3880 3890 3900
KC955130_BG8 TGGGTGAGTCTTCTTTCCCCAACCCCAAGAAATATGCGTTTCCCATGGGATGACAAGCTGTGCCACCTCATCATGCCCTGTTTTTTCTGTCCTTTTTGCA
NBBGa-1(24,2) ...
NBBGa-2(15,1) ...
NBBGa-3(5,T) ...
NBBGa-4(4,2) ...
NBBGa-5(2,1) ...
NBBGa-6(2,1) ...
NBBGa-7(2,T) ...
NBBGa-8(2,1) ...
NBBGa-9(2,1) ...
NBBGa-10(1,1) ...
NBBGa-11(1,1) ...
NBBGa-13(1,1) ...
NBBGa-14(1,1) ...
NBBGa-15(1,1) ...
NBBGa-17(1,1) ...
NBBGa-18(1,1) ...
NBBGa-21(1,1) ...
NBBGa-22(27,V)
NBBGf-1(1,1) ...
NBBGf-2(1,1) ...
NBBGf-3(10,V)
3910 3920 3930 3940 3950 3960 3970 3980 3990 4000
KC955130_BG8 GAGAAACAGCATTCACAGTTCCGTAAGTTGCAGTCACTGAACTGAAGGAATGTGGGGTCTTCCCAAAGTCCTGCATGTGGGATG AAAAATCCCCTCTGA
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1) .....A.....AC.....A.....
NBBGf-3(10,V)

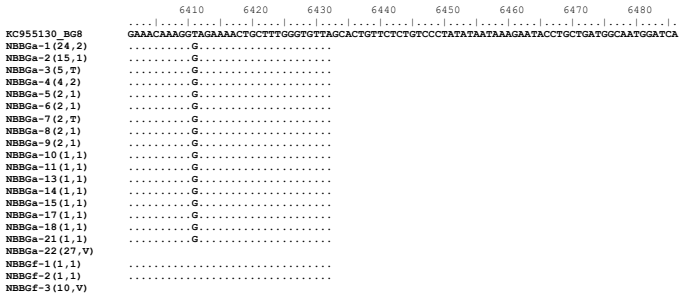

4010 4020 4030 4040 4050 4060 4070 4080 4090 4100
KC955130_BG8 CCATGCACTGCTTTTCTCCTTCTATTCCAGAGAGACACTTTCAGAATATGTGTGAGTCTCCCCACCCCTGATAAATAAAAACGTTGGGGGTCTTGCTGTGT
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1) .....CG....G
NBBGf-2(1,1) .....T.....C.....CG....G
NBBGf-3(10,V)
4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
KC955130_BG8 GAGCTGTGGGATGAGATGTTCCTCTCATCACACATTGTTTTTCTGATTTTCTTTTGCAGATTTAAGTGCTGGAAAACAGAGTAAGTCTCCCTCCCTGCAC
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1) .....T...C....
NBBGf-2(1,1) .....T...C....
NBBGf-3(10,V)
4210 4220 4230 4240 4250 4260 4270 4280 4290 4300
KC955130_BG8 AGAAGGAACTTACGGTTTCCCATGGGATCAGCCATGGGATCATCATCCGACTCTTCTCATCATGAATTTCGTCTTTCTTTCTTTTGCAGAGAAAATGGT
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1) .....CG.....
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1) .....C....
NBBGf-2(1,1) .....C....
NBBGf-3(10,V)
4310 4320 4330 4340 4350 4360 4370 4380 4390 4400
KC955130_BG8 TACAAAACTGGGTGAGTCCAACCTCCCAAACTAAATTAAAAACAGTCAGACTTTGTGAGCTGTGGGATGAGACGTTCCTCTCATCATGTGCTGCTTTCCT
NBBGa-1(24,2) ....
NBBGa-2(15,1) ....
NBBGa-3(5,T) ....
NBBGa-4(4,2) ....
NBBGa-5(2,1) ....
NBBGa-6(2,1) ....
NBBGa-7(2,T) ....
NBBGa-8(2,1) ....
NBBGa-9(2,1) ....
NBBGa-10(1,1) ....
NBBGa-11(1,1) ....
NBBGa-13(1,1) ....
NBBGa-14(1,1) ....
NBBGa-15(1,1) ....
NBBGa-17(1,1) ....
NBBGa-18(1,1) ....
NBBGa-21(1,1) ....
NBBGa-22(27,V)
NBBGf-1(1,1) ....
NBBGf-2(1,1) ....
NBBGf-3(10,V)


4410 4420 4430 4440 4450 4460 4470 4480 4490 4500
KC955130_BG8 TTTACTTTTCCAGAGGAACACTGTGAATGGATGGGTGAGTCTCCCCTCCCAAATTAAAAATGTTGGGGGTCTTCCTGTGAGAGCTGTGGGATGAGCTGTTC
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1) .....T.....
NBBGf-2(1,1) .....T.....
NBBGf-3(10,V)
4510 4520 4530 4540 4550 4560 4570 4580 4590 4600
KC955130_BG8 CTCTCATCGTGCACTGTTTCTGCTTTCCTTTGCAGTGAGAAGGAATGTAAAGTTGGGTGAGTCTTCTTCCCCAACCAAAGAGATTCGGAGTCTTCCATG
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1) A.....
NBBGf-2(1,1) A.....
NBBGf-3(10,V)
4610 4620 4630 4640 4650 4660 4670 4680 4690 4700
KC955130_BG8 GGATCAGCCATGGGATGATAACATGAACCTCATCACGTGTTTCTTATTTGTTCCTTTTGCAGAGCCAGCAGCTGTAAAAGTGGGTGAGTCTCCCCTCCCA
NBBGa-1(24,2) ..AT.C.....
NBBGa-2(15,1) ..AT.C.....
NBBGa-3(5,T) ..AT.C.....
NBBGa-4(4,2) ..AT.C.....
NBBGa-5(2,1) ..AT.C.....
NBBGa-6(2,1) ..AT.C.....
NBBGa-7(2,T) ..AT.C.....
NBBGa-8(2,1) ..AT.C.....
NBBGa-9(2,1) ..AT.C.....
NBBGa-10(1,1) ..AT.C.....
NBBGa-11(1,1) ..AT.C.....
NBBGa-13(1,1) ..AT.C.....
NBBGa-14(1,1) ..AT.C.....
NBBGa-15(1,1) ..AT.C.....
NBBGa-17(1,1) ..AT.C.....
NBBGa-18(1,1) ..AT.C.....
NBBGa-21(1,1) ..AT.C.....
NBBGa-22(27,V)
NBBGf-1(1,1) ..AT.C.....T.C...
NBBGf-2(1,1) ..AT.C.....T.C...
NBBGf-3(10,V)
4710 4720 4730 4740 4750 4760 4770 4780 4790 4800
KC955130_BG8 AATTAAAAATGTTGGCGTCATCCTGTGAGAGCTGTGGGATGAGCTGTTCCTCTCATCGTGCACTGTTTCTGCTTTTCCTTTGCAGTGAGAAGGAATGTAA
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)

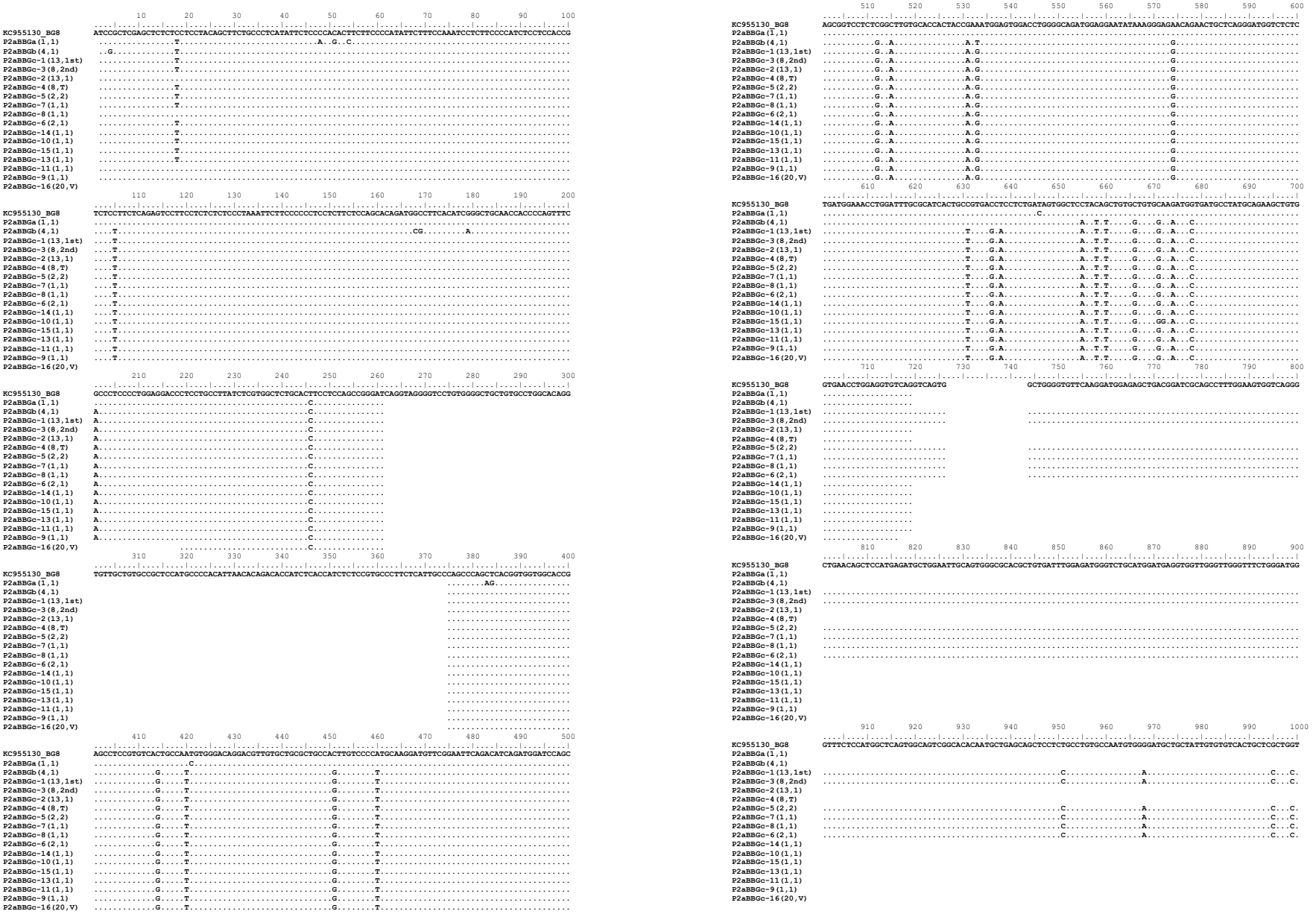

4810 4820 4830 4840 4850 4860 4870 4880 4890 4900
KC955130_BG8 AGTTGGGTGAGTCTTCTTCCCCAACCAAAGAGATGTGGGGTCTTCCATGGGATCAGCCATGGGATGATAAGCTGAACCTTATCACGTGTTTCTTATTTGT
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
4910 4920 4930 4940 4950 4960 4970 4980 4990 5000
KC955130_BG8 TCCTTTTGCAGAGGCAGCAGCTGTAAAAGTGGGTGAGTCCTCCCTCCCAAATCAAATACAAAAGGGGATCTGCCTGTGTGAGCTGTGGGATGAGATGTTC
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
5010 5020 5030 5040 5050 5060 5070 5080 5090 5100
KC955130_BG8 CTCTCATCACGCATTGTTTTTCTCATTCATTTCCAGGACACAAAGCTAAAGAATCAGGTGAGTCTTCTTCCCTGTCCCAAAGGACTATGGGTTTCCCATG
NBBGa-1(24,2) .G..AC.....
NBBGa-2(15,1) .G..AC.....
NBBGa-3(5,T) .G..AC.....
NBBGa-4(4,2) .G..AC.....
NBBGa-5(2,1) .G..AC.....
NBBGa-6(2,1) .G..AC.....
NBBGa-7(2,T) .G..AC.....
NBBGa-8(2,1) ....AC.....
NBBGa-9(2,1) .G..AC.....
NBBGa-10(1,1) .G..AC.....
NBBGa-11(1,1) .G..AC.....
NBBGa-13(1,1) .G..AC.....
NBBGa-14(1,1) .G..AC.....
NBBGa-15(1,1) .G..AC.....
NBBGa-17(1,1) .G..AC.....
NBBGa-18(1,1) .G..AC.....
NBBGa-21(1,1) .G..AC.....
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
5110 5120 5130 5140 5150 5160 5170 5180 5190 5200
KC955130_BG8 GGATGACAAGCTGTGCCACCTCCTCATGAGGTGCTTCTTCTTTCTTTGTGCAGAGAAACAGAAATCGGAGCTGAGTAAGTTGCAGTCACTGAACTGAGGG
NBBGa-1(24,2) ..G.....
NBBGa-2(15,1) ..G.....
NBBGa-3(5,T) ..G.....
NBBGa-4(4,2) ..G.....
NBBGa-5(2,1) ..G.....
NBBGa-6(2,1) ..G.....
NBBGa-7(2,T) ..G.....
NBBGa-8(2,1) ..G.....
NBBGa-9(2,1) ..G.....
NBBGa-10(1,1) ..G.....
NBBGa-11(1,1) ..G.....
NBBGa-13(1,1) ..G.....
NBBGa-14(1,1) ..G.....
NBBGa-15(1,1) ..G.....
NBBGa-17(1,1) ..G.....
NBBGa-18(1,1) ..G.....
NBBGa-21(1,1) ..G.....
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
5210 5220 5230 5240 5250 5260 5270 5280 5290 5300
KC955130_BG8 AATGTGGGGTCTTCCCAAAGTCCTGCGTATGGGATGAAAAATCCCCTCTGACCATGCACTGCTTTTCTCCTCCTTTGCCAGAGGAGCGCCATGAGCAGAT
NBBGa-1(24,2) ......A.....C
NBBGa-2(15,1) ......A.....C
NBBGa-3(5,T) ......A.....C
NBBGa-4(4,2) ......A.....C
NBBGa-5(2,1) ......A.....C
NBBGa-6(2,1) ......A.....C
NBBGa-7(2,T) ......A.....C
NBBGa-8(2,1) ......A.....C
NBBGa-9(2,1) ......A.....C
NBBGa-10(1,1) ......A.....C
NBBGa-11(1,1) ......A.....C
NBBGa-13(1,1) ......A.....C
NBBGa-14(1,1) ......A.....C
NBBGa-15(1,1) ......A.....C
NBBGa-17(1,1) ......A.....C
NBBGa-18(1,1) ......A.....C
NBBGa-21(1,1) ......A.....C
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
5310 5320 5330 5340 5350 5360 5370 5380 5390 5400
KC955130_BG8 GGGTGAGTCTCCCCTCCCATATTAAAATCGTTGGGGTCTTCCTGTGTGAGCTGTGGGATGAGATGTTCCTCTCATCACACATTGTTTTTCTTTTCCAGGG
NBBGa-1(24,2) ..
NBBGa-2(15,1) ..
NBBGa-3(5,T) ..
NBBGa-4(4,2) ..
NBBGa-5(2,1) ..
NBBGa-6(2,1) ..
NBBGa-7(2,T) ..
NBBGa-8(2,1) ..
NBBGa-9(2,1) ..
NBBGa-10(1,1) ..
NBBGa-11(1,1) ..
NBBGa-13(1,1) ..
NBBGa-14(1,1) ..
NBBGa-15(1,1) ..
NBBGa-17(1,1) ..
NBBGa-18(1,1) ..
NBBGa-21(1,1) ..
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
5410 5420 5430 5440 5450 5460 5470 5480 5490 5500
KC955130_BG8 CAACAAGCTAAAGAATCAGGTGAGTCTTCTTCCCGTCCCAAAGGACTATGGGTTTCCCATGGGATGACAAGCTGTGCCACCTCCTCATGAGGTGCTTCT
NBBGa-1(24,2) ....
NBBGa-2(15,1) ....
NBBGa-3(5,T) ....
NBBGa-4(4,2) ....
NBBGa-5(2,1) ....
NBBGa-6(2,1) ....
NBBGa-7(2,T) ....
NBBGa-8(2,1) ....
NBBGa-9(2,1) ....
NBBGa-10(1,1) ....
NBBGa-11(1,1) ....
NBBGa-13(1,1) ....
NBBGa-14(1,1) ....
NBBGa-15(1,1) ....
NBBGa-17(1,1) ....
NBBGa-18(1,1) ....
NBBGa-21(1,1) ....
NBBGa-22(27,V)
NBBGf-1(1,1) ....
NBBGf-2(1,1) ....
NBBGf-3(10,V)
5510 5520 5530 5540 5550 5560 5570 5580 5590 5600
KC955130_BG8 TCTTTCTTTTTTGCAGAGAAACAGAAATCGGAGCTGAGTAAGTTGCAGTCACTGAACTGAGGGTATTTGGGGTCCTTTCAAGGGACTGTGTATGGGATGA
NBBGa-1(24,2) ....
NBBGa-2(15,1) ....
NBBGa-3(5,T) ....
NBBGa-4(4,2) ....
NBBGa-5(2,1) ....
NBBGa-6(2,1) ....
NBBGa-7(2,T) ....
NBBGa-8(2,1) ....
NBBGa-9(2,1) ....
NBBGa-10(1,1) ....
NBBGa-11(1,1) ....
NBBGa-13(1,1) ....
NBBGa-14(1,1) ....
NBBGa-15(1,1) ....
NBBGa-17(1,1) ....
NBBGa-18(1,1) ....
NBBGa-21(1,1) ....
NBBGa-22(27,V)
NBBGf-1(1,1) ....
NBBGf-2(1,1) ....
NBBGf-3(10,V)

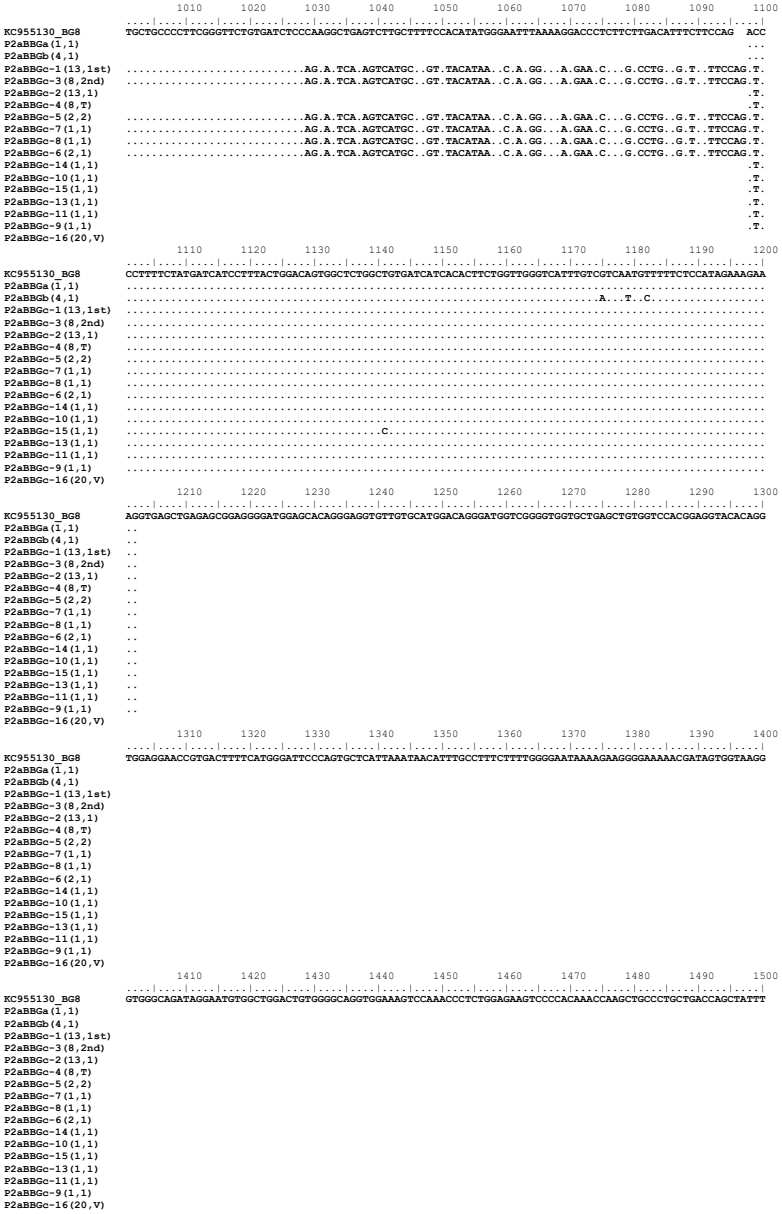

5610 5620 5630 5640 5650 5660 5670 5680 5690 5700
KC955130_BG8 AAAATCCCCTCTGACCATGCACTGCTTTTCTCCTTCTTTGCCAGAGGAGCGCCCATGAGGAGATGGGTGAGTCTCCCCTCCCATATTAAAATCGTTGGGGT
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
5710 5720 5730 5740 5750 5760 5770 5780 5790 5800
KC955130_BG8 CTTCCTGTGTGAGCTGTGGGATGAGATGTTCCTCTCATCGTGTGGTGCTTTTCTCTCTTTTCCAGCAGAACAAACTGAAGCAGTGGGTGAGTCTTTGTCC
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1)
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
5810 5820 5830 5840 5850 5860 5870 5880 5890 5900
KC955130_BG8 CCAAACCAAAGGAATATGGGGCAATCCATGGGATGACAAGCTGTCCCATCTCATCATGTGTTGCTTTTCTATTTCCTTTTTT CCAGTGGTAGAAACTGAA
NBBGa-1(24,2)
NBBGa-2(15,1)
NBBGa-3(5,T)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-6(2,1)
NBBGa-7(2,T)
NBBGa-8(2,1)
NBBGa-9(2,1)
NBBGa-10(1,1)
NBBGa-11(1,1) ...G.....G...A.....G.....T.....
NBBGa-13(1,1)
NBBGa-14(1,1)
NBBGa-15(1,1)
NBBGa-17(1,1) ...G.....G...A.....G.....
NBBGa-18(1,1) ...G.....G...A.....G.....
NBBGa-21(1,1)
NBBGa-22(27,V)
NBBGf-1(1,1) .....T.....
NBBGf-2(1,1) .....T.....
NBBGf-3(10,V)
5910 5920 5930 5940 5950 5960 5970 5980 5990 6000
KC955130_BG8 GAATAGGGTGAGTCTTTCCCAAACCAAAGCAATACAGGGTTTCCCATGGGATGACAAGCTGTCCCACCTCAGCATCCGTTCCTTTTTATTTCTTTCCAG
NBBGa-1(24,2) ....
NBBGa-2(15,1) ....
NBBGa-3(5,T) ....
NBBGa-4(4,2) ....
NBBGa-5(2,1) ....
NBBGa-6(2,1) ....
NBBGa-7(2,T) .....G.....C.....C.....
NBBGa-8(2,1) .....G.....C.....C.....
NBBGa-9(2,1) .....G.....C.....C.....
NBBGa-10(1,1) ....
NBBGa-11(1,1) ....
NBBGa-13(1,1) ....
NBBGa-14(1,1) ....
NBBGa-15(1,1) ....
NBBGa-17(1,1) .....G.....C.....C.....
NBBGa-18(1,1) .....G.....C.....C.....
NBBGa-21(1,1) .....G.....C.....C.....
NBBGa-22(27,V)
NBBGf-1(1,1) ...GC..
NBBGf-2(1,1) ...GC..
NBBGf-3(10,V)

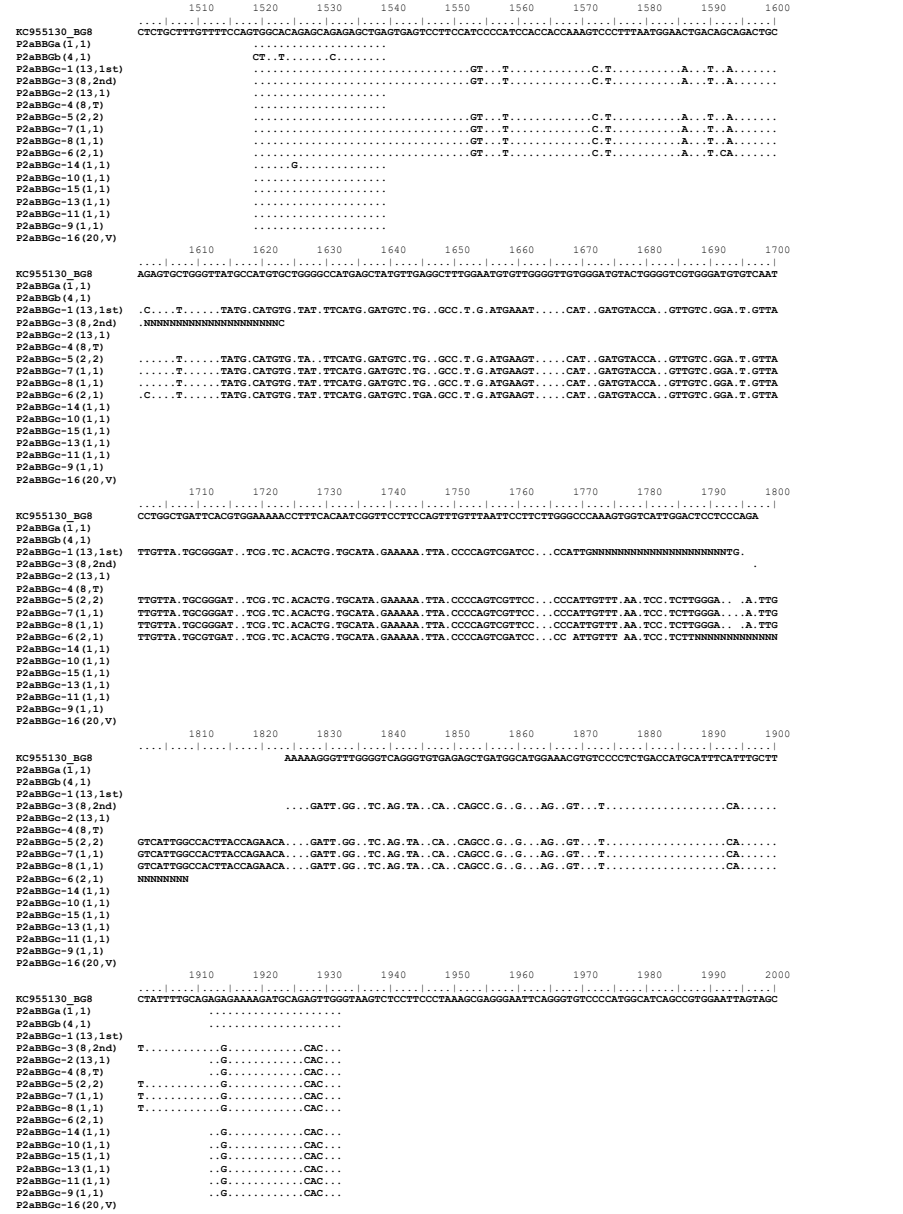

6010 6020 6030 6040 6050 6060 6070 6080 6090 6100
KC955130_BG8 AAAAACCATCTGAAGAATCAGATTGAGAGATGAACTGCGCCTCGCAAATAAGCACAGGAGTTAAGCTTCATAGATCAATAACTGCACAGCATAC AAAACC
NBBGa-1(24,2) .....TG.....G..C.....
NBBGa-2(15,1) .....TG.....G..C.....
NBBGa-3(5,T) .....TG.....G..C.....
NBBGa-4(4,2) .....TG.....G..C.....
NBBGa-5(2,1) .....TG.....G..C.....
NBBGa-6(2,1) .....TG.....G..C.....
NBBGa-7(2,T) .....TG.....G..C.....
NBBGa-8(2,1) .....TG.....G..C.....
NBBGa-9(2,1) .....TG.....G..C.....
NBBGa-10(1,1) .....TG.....G..C.....
NBBGa-11(1,1) .....TG.....G..C.....
NBBGa-13(1,1) .....TG.....G..C.....
NBBGa-14(1,1) .....TG.....G..C.....
NBBGa-15(1,1) .....TG.....G..C.....
NBBGa-17(1,1) .....TG.....G..C.....
NBBGa-18(1,1) .....TG.....G..C.....
NBBGa-21(1,1) .....TG.....G..C.....
NBBGa-22(27,V)
NBBGf-1(1,1) .....TG.....A..G..C.....G.....A.....
NBBGf-2(1,1) .....TG.....A..G..C.....G.....A.....
NBBGf-3(10,V)
6110 6120 6130 6140 6150 6160 6170 6180 6190 6200
KC955130_BG8 ACAATAACTCAAACAGAGTAAGGA GGAGCCAGTGTTTGTGTTGAGTGAGAACACTGCAGTTCTGTCAGCCAAAGCTGCCT
NBBGa-1(24,2) .....G.....
NBBGa-2(15,1) .....G.....
NBBGa-3(5,T) .....G.....
NBBGa-4(4,2) .....G.....
NBBGa-5(2,1) .....G.....
NBBGa-6(2,1) .....G.....
NBBGa-7(2,T) .....G.....
NBBGa-8(2,1) .....G.....
NBBGa-9(2,1) .....G.....
NBBGa-10(1,1) .....G.....
NBBGa-11(1,1) .....G.....
NBBGa-13(1,1) .....G.....
NBBGa-14(1,1) .....G.....
NBBGa-15(1,1) .....G.....
NBBGa-17(1,1) .....G.....
NBBGa-18(1,1) .....G.....
NBBGa-21(1,1) .....G.....
NBBGa-22(27,V)
NBBGf-1(1,1) ..G..C.....C.....AATCCACAGCGAGAACAAGA.....
NBBGf-2(1,1) ..G..C.....C.....AATCCACAGCGAGAACAAGA.....
NBBGf-3(10,V)
6210 6220 6230 6240 6250 6260 6270 6280 6290 6300
KC955130_BG8 GAGGGACCGCCGAATTGAGGGTGTGCGACCTCCAACTCAAAGCCAATTGGAAGAAAGAAACCATAGAAAGGAAGCAAAGGGGAGGGAGACAGAGATCCTG
NBBGa-1(24,2) .....C.....A.....A.....
NBBGa-2(15,1) .....C.....A.....A.....
NBBGa-3(5,T) .....C.....A.....A.....
NBBGa-4(4,2) .....C.....A.....A.....
NBBGa-5(2,1) .....C.....A.....A.....
NBBGa-6(2,1) .....C.....A.....A.....
NBBGa-7(2,T) .....C.....A.....A.....
NBBGa-8(2,1) .....C.....A.....A.....
NBBGa-9(2,1) .....C.....A.....A.....
NBBGa-10(1,1) .....C.....A.....A.....
NBBGa-11(1,1) .....C.....A.....A.....
NBBGa-13(1,1) .....C.....A.....A.....
NBBGa-14(1,1) .....C.....A.....A.....
NBBGa-15(1,1) .....C.....A.....A.....
NBBGa-17(1,1) .....C.....A.....A.....
NBBGa-18(1,1) .....C.....A.....A.....
NBBGa-21(1,1) .....C.....A.....A.....
NBBGa-22(27,V)
NBBGf-1(1,1) .....A..C.....A.....
NBBGf-2(1,1) .....A..C.....A.....
NBBGf-3(10,V)
6310 6320 6330 6340 6350 6360 6370 6380 6390 6400
KC955130_BG8 GAAGAGATATGGGCATTTGGGGAAATAGTGTGACCGTGTATCAGGCTTTGTGGACATCTAACGAATATGTCATGTTTTTGTAAATACAAGCATGCACGCA
NBBGa-1(24,2) .....A.....
NBBGa-2(15,1) .....A.....
NBBGa-3(5,T) .....A.....
NBBGa-4(4,2) .....A.....
NBBGa-5(2,1) .....A.....
NBBGa-6(2,1) .....A.....
NBBGa-7(2,T) .....A.....
NBBGa-8(2,1) .....A.....
NBBGa-9(2,1) .....A.....
NBBGa-10(1,1) .....A.....
NBBGa-11(1,1) .....A.....
NBBGa-13(1,1) .....A.....A.....G.....
NBBGa-14(1,1) .....A.....A.....
NBBGa-15(1,1) .....A.....G.....G.....
NBBGa-17(1,1) .....A.....
NBBGa-18(1,1) .....A.....
NBBGa-21(1,1) .....A.....
NBBGa-22(27,V)
NBBGf-1(1,1) .....A.....G.....T..
NBBGf-2(1,1) .....A.....G.....T..
NBBGf-3(10,V)

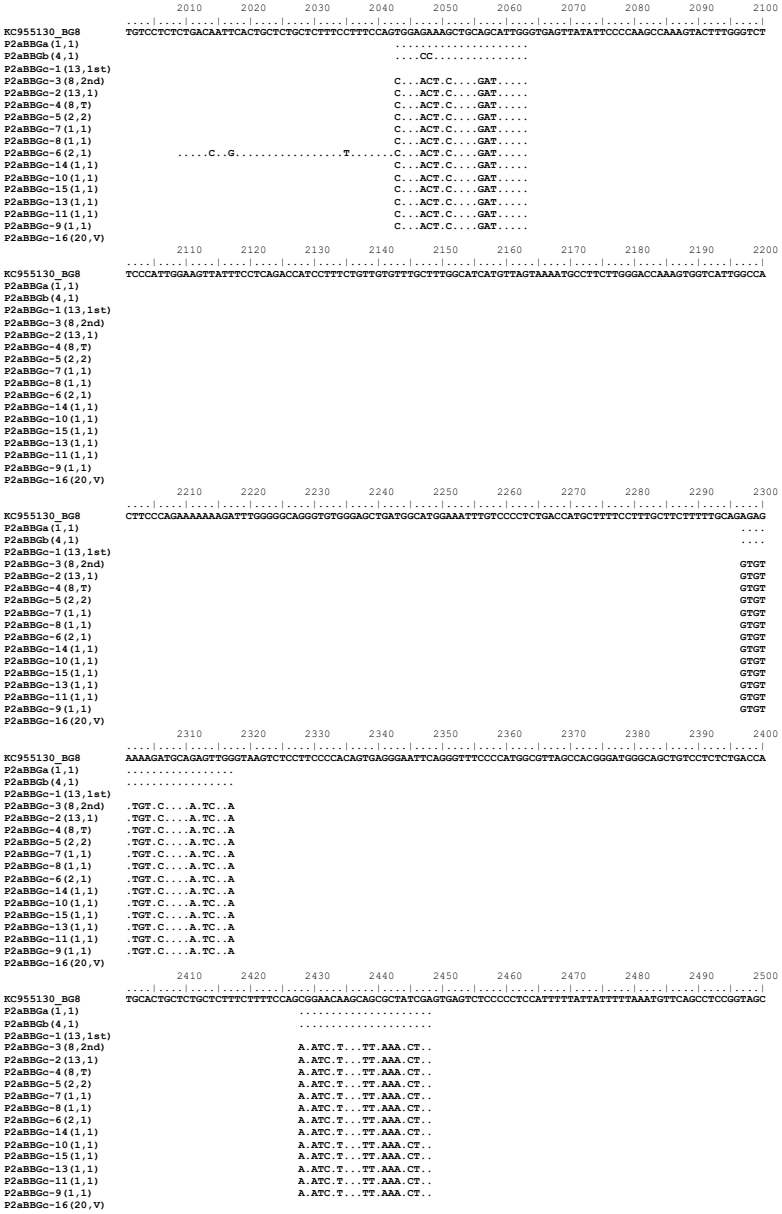

6410 6420 6430 6440 6450 6460 6470 6480
KC955130_BG8 GAAACAAAGGTAGAAACTGCTTTGGGTGTTAGCACTGTTCTCTGTCCCTATATAATAAAGAATACCTGCTGATGGCAATGGATCA
NBBGa-1(24,2) ....G.....................
NBBGa-2(15,1) ....G.....................
NBBGa-3(5,T) ....G.....................
NBBGa-4(4,2) ....G.....................
NBBGa-5(2,1) ....G.....................
NBBGa-6(2,1) ....G.....................
NBBGa-7(2,T) ....G.....................
NBBGa-8(2,1) ....G.....................
NBBGa-9(2,1) ....G.....................
NBBGa-10(1,1) ....G.....................
NBBGa-11(1,1) ....G.....................
NBBGa-13(1,1) ....G.....................
NBBGa-14(1,1) ....G.....................
NBBGa-15(1,1) ....G.....................
NBBGa-17(1,1) ....G.....................
NBBGa-18(1,1) ....G.....................
NBBGa-21(1,1) ....G.....................
NBBGa-22(27,V)
NBBGf-1(1,1) ....
NBBGf-2(1,1) ....
NBBGf-3(10,V)

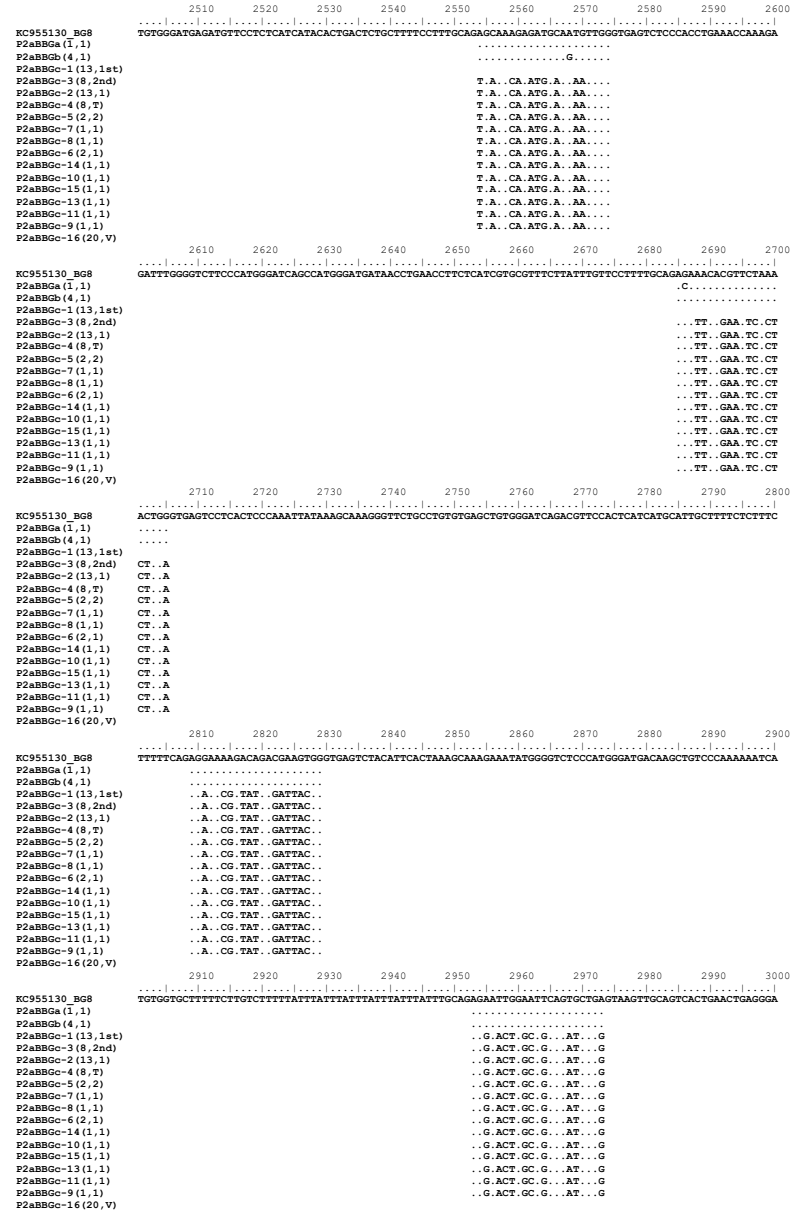
(J2)

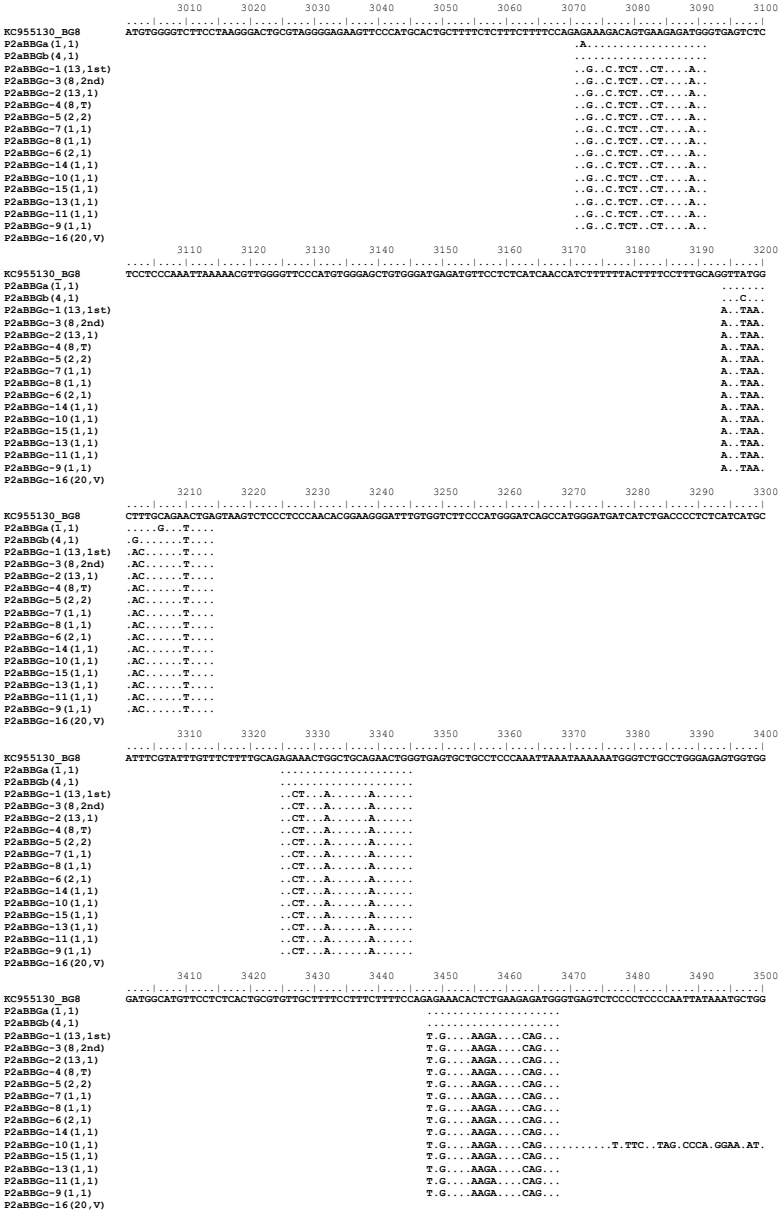

KC955130_BG8
P2aBBGa(1,1)
P2aBBGb(4,1)
P2aBBGc-1(13,1st)
P2aBBGc-3(8,2nd)
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2)
P2aBBGc-7(1,1)
P2aBBGc-8(1,1)
P2aBBGc-6(2,1)
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)

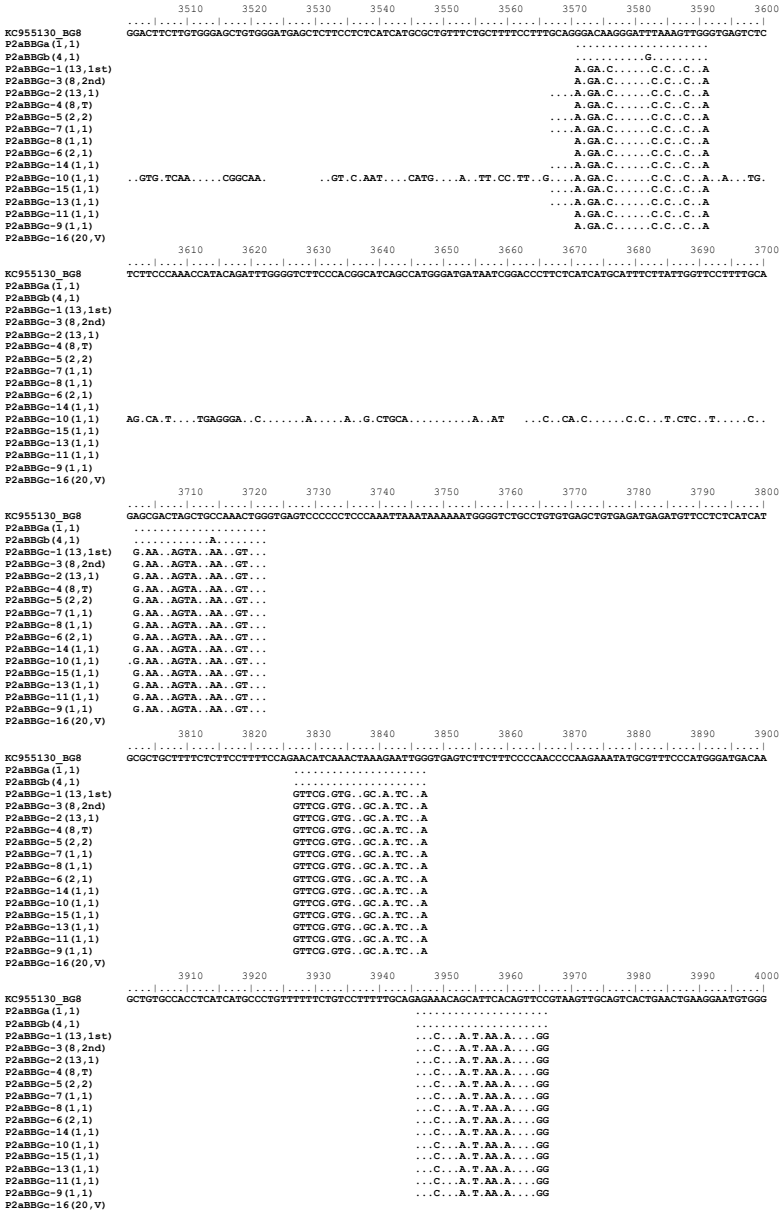

1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
KC955130_BG8 TGCTGCCCCTTCGGGTTCTGTGATCTCCCAAGGCTGAGTCTTGCTTTTCCACATATGGGAATTTAAAAGGACCCTCTTCTTGACATTTCTTCCAG ACC
P2aBBGa(1,1) ...
P2aBBGb(4,1) ...
P2aBBGc-1(13,1st) .....AG.A.TCA.AGTCATGC..GT.TACATAA..C.A.GG...A.GAA.C...G.CCTG..G.T..TTCCAG.T.
P2aBBGc-3(8,2nd) .....AG.A.TCA.AGTCATGC..GT.TACATAA..C.A.GG...A.GAA.C...G.CCTG..G.T..TTCCAG.T.
P2aBBGc-2(13,1) .T.
P2aBBGc-4(8,T) .T.
P2aBBGc-5(2,2) .....AG.A.TCA.AGTCATGC..GT.TACATAA..C.A.GG...A.GAA.C...G.CCTG..G.T..TTCCAG.T.
P2aBBGc-7(1,1) .....AG.A.TCA.AGTCATGC..GT.TACATAA..C.A.GG...A.GAA.C...G.CCTG..G.T..TTCCAG.T.
P2aBBGc-8(1,1) .....AG.A.TCA.AGTCATGC..GT.TACATAA..C.A.GG...A.GAA.C...G.CCTG..G.T..TTCCAG.T.
P2aBBGc-6(2,1) .....AG.A.TCA.AGTCATGC..GT.TACATAA..C.A.GG...A.GAA.C...G.CCTG..G.T..TTCCAG.T.
P2aBBGc-14(1,1) .T.
P2aBBGc-10(1,1) .T.
P2aBBGc-15(1,1) .T.
P2aBBGc-13(1,1) .T.
P2aBBGc-11(1,1) .T.
P2aBBGc-9(1,1) .T.
P2aBBGc-16(20,V)
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
KC955130_BG8 CCTTTTCTATGATCATCCTTTACTGGACAGTGGCTCTGGCTGTGATCATCACACTTCTGGTTGGGTCATTTGTCGTCAATGTTTTTCTCCATAGAAAAGAA
P2aBBGa(1,1) ...
P2aBBGb(4,1) ...A..T..C...
P2aBBGc-1(13,1st) ...
P2aBBGc-3(8,2nd) ...
P2aBBGc-2(13,1) ...
P2aBBGc-4(8,T) ...
P2aBBGc-5(2,2) ...
P2aBBGc-7(1,1) ...
P2aBBGc-8(1,1) ...
P2aBBGc-6(2,1) ...
P2aBBGc-14(1,1) ...
P2aBBGc-10(1,1) ...
P2aBBGc-15(1,1) ...C...
P2aBBGc-13(1,1) ...
P2aBBGc-11(1,1) ...
P2aBBGc-9(1,1) ...
P2aBBGc-16(20,V)
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
KC955130_BG8 AGGTGAGCTGAGAGCGGAGGGGATGGAGCACAGGGAGGTGTTGTGCATGGACAGGGATGGTCGGGGTGGTGCTGAGCTGTGGTCCACGGAGGTACACAGG
P2aBBGa(1,1) ..
P2aBBGb(4,1) ..
P2aBBGc-1(13,1st) ..
P2aBBGc-3(8,2nd) ..
P2aBBGc-2(13,1) ..
P2aBBGc-4(8,T) ..
P2aBBGc-5(2,2) ..
P2aBBGc-7(1,1) ..
P2aBBGc-8(1,1) ..
P2aBBGc-6(2,1) ..
P2aBBGc-14(1,1) ..
P2aBBGc-10(1,1) ..
P2aBBGc-15(1,1) ..
P2aBBGc-13(1,1) ..
P2aBBGc-11(1,1) ..
P2aBBGc-9(1,1) ..
P2aBBGc-16(20,V)
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
KC955130_BG8 TGGAGGAACCGTGACTTTTCATGGGATTCCCAGTGCTCATTAAATAACATTTGCCTTTCTTTTGGGGAATAAAAGAAGGGGAAAAACGATAGTGGTAAGG
P2aBBGa(1,1) ...
P2aBBGb(4,1) ...
P2aBBGc-1(13,1st) ...
P2aBBGc-3(8,2nd) ...
P2aBBGc-2(13,1) ...
P2aBBGc-4(8,T) ...
P2aBBGc-5(2,2) ...
P2aBBGc-7(1,1) ...
P2aBBGc-8(1,1) ...
P2aBBGc-6(2,1) ...
P2aBBGc-14(1,1) ...
P2aBBGc-10(1,1) ...
P2aBBGc-15(1,1) ...
P2aBBGc-13(1,1) ...
P2aBBGc-11(1,1) ...
P2aBBGc-9(1,1) ...
P2aBBGc-16(20,V)
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
KC955130_BG8 GTGGGCAGATAGGAATGTGGCTGGACTGTGGGGCAGGTGGAAAGTCCAAACCCTCTGGAGAAGTCCCCACAAACCAAGCTGCCCTGCTGACCAGCTATTT
P2aBBGa(1,1) ...
P2aBBGb(4,1) ...
P2aBBGc-1(13,1st) ...
P2aBBGc-3(8,2nd) ...
P2aBBGc-2(13,1) ...
P2aBBGc-4(8,T) ...
P2aBBGc-5(2,2) ...
P2aBBGc-7(1,1) ...
P2aBBGc-8(1,1) ...
P2aBBGc-6(2,1) ...
P2aBBGc-14(1,1) ...
P2aBBGc-10(1,1) ...
P2aBBGc-15(1,1) ...
P2aBBGc-13(1,1) ...
P2aBBGc-11(1,1) ...
P2aBBGc-9(1,1) ...
P2aBBGc-16(20,V)

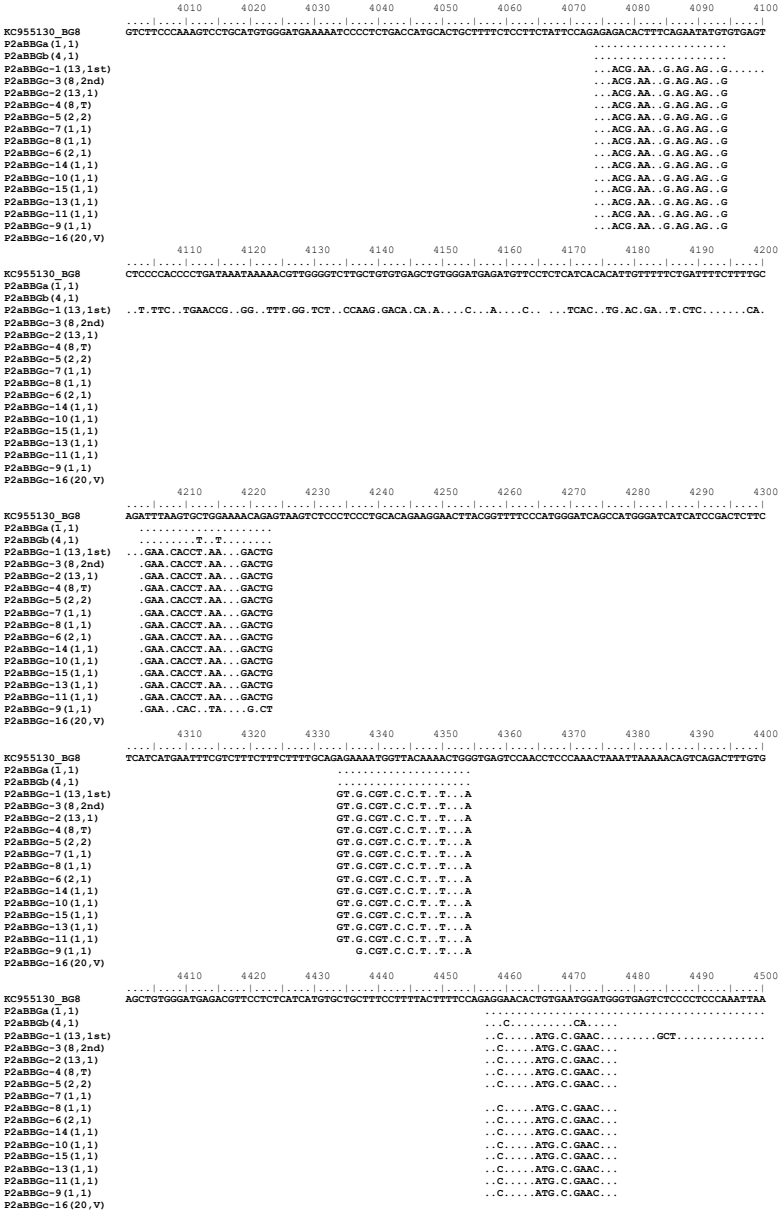

1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
KC955130_BG8 CTCTGCTTTGTTTTCCAGTGGCACAGAGCAGAGAGCTGAGTGAGTCCTTCCATCCCCATCCACCACCAAAGTCCCTTTAATGGAACTGACAGCAGACTGC
P2aBBGa(1,1) .....
P2aBBGb(4,1) CT..T.....C.....
P2aBBGc-1(13,1st) .....GT...T.....C.T.....A...T..A.....
P2aBBGc-3(8,2nd) .....GT...T.....C.T.....A...T..A.....
P2aBBGc-2(13,1) .....
P2aBBGc-4(8,T) .....
P2aBBGc-5(2,2) .....GT...T.....C.T.....A...T..A.....
P2aBBGc-7(1,1) .....GT...T.....C.T.....A...T..A.....
P2aBBGc-8(1,1) .....GT...T.....C.T.....A...T..A.....
P2aBBGc-6(2,1) .....GT...T.....C.T.....A...T.CA.....
P2aBBGc-14(1,1) ......G.....
P2aBBGc-10(1,1) .....
P2aBBGc-15(1,1) .....
P2aBBGc-13(1,1) .....
P2aBBGc-11(1,1) .....
P2aBBGc-9(1,1) .....
P2aBBGc-16(20,V)
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
KC955130_BG8 AGAGTGCTGGGTTATGCCATGTGCTGGGGCCATGAGCTATGTTGAGGCTTTGGAATGTGTTGGGGTTGTGGGATGTACTGGGGTCGTGGGATGTGTCAAT
P2aBBGa(1,1)
P2aBBGb(4,1)
P2aBBGc-1(13,1st) .C....T......TATG.CATGTG.TAT.TTCATG.GATGTC.TG..GCC.T.G.ATGAAAT.....CAT..GATGTACCA..GTTGTC.GGA.T.GTTA
P2aBBGc-3(8,2nd) .NNNNNNNNNNNNNNNNNNNNC
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2) ......T......TATG.CATGTG.TA..TTCATG.GATGTC.TG..GCC.T.G.ATGAAGT.....CAT..GATGTACCA..GTTGTC.GGA.T.GTTA
P2aBBGc-7(1,1) ......T......TATG.CATGTG.TAT.TTCATG.GATGTC.TG..GCC.T.G.ATGAAGT.....CAT..GATGTACCA..GTTGTC.GGA.T.GTTA
P2aBBGc-8(1,1) ......T......TATG.CATGTG.TAT.TTCATG.GATGTC.TG..GCC.T.G.ATGAAGT.....CAT..GATGTACCA..GTTGTC.GGA.T.GTTA
P2aBBGc-6(2,1) .C....T......TATG.CATGTG.TAT.TTCATG.GATGTC.TGA.GCC.T.G.ATGAAGT.....CAT..GATGTACCA..GTTGTC.GGA.T.GTTA
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
KC955130_BG8 CCTGGCTGATTCACGTGGAAAAACCTTTCACAATCGGTTCCTTCCAGTTTGTTTAATTCCTTCTTGGGCCCAAAGTGGTCATTGGACTCCTCCCAGA
P2aBBGa(1,1)
P2aBBGb(4,1)
P2aBBGc-1(13,1st) TTGTTA.TGCGGGAT..TCG.TC.ACACTG.TGCATA.GAAAAA.TTA.CCCCAGTCGATCC...CCATTGNNNNNNNNNNNNNNNNNNNNTG.
P2aBBGc-3(8,2nd) .
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2) TTGTTA.TGCGGGAT..TCG.TC.ACACTG.TGCATA.GAAAAA.TTA.CCCCAGTCGTTCC...CCCATTGTTT.AA.TCC.TCTTGGGA.. .A.TTG
P2aBBGc-7(1,1) TTGTTA.TGCGGGAT..TCG.TC.ACACTG.TGCATA.GAAAAA.TTA.CCCCAGTCGTTCC...CCCATTGTTT.AA.TCC.TCTTGGGA...A.TTG
P2aBBGc-8(1,1) TTGTTA.TGCGGGAT..TCG.TC.ACACTG.TGCATA.GAAAAA.TTA.CCCCAGTCGTTCC...CCCATTGTTT.AA.TCC.TCTTGGGA.. .A.TTG
P2aBBGc-6(2,1) TTGTTA.TGCGTGAT..TCG.TC.ACACTG.TGCATA.GAAAAA.TTA.CCCCAGTCGATCC...CC ATTGTTT AA.TCC.TCTTNNNNNNNNNNNNN
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
KC955130_BG8 AAAAAGGGTTTGGGGTCAGGGTGTGAGAGCTGATGGCATGGAAACGTGTCCCCTCTGACCATGCATTTCATTTGCTT
P2aBBGa(1,1)
P2aBBGb(4,1)
P2aBBGc-1(13,1st)
P2aBBGc-3(8,2nd) ....GATT.GG..TC.AG.TA..CA..CAGCC.G..G...AG..GT...T.....CA.....
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2) GTCATTGGCCACTTACCAGAACA....GATT.GG..TC.AG.TA..CA..CAGCC.G..G...AG..GT...T.....CA.....
P2aBBGc-7(1,1) GTCATTGGCCACTTACCAGAACA....GATT.GG..TC.AG.TA..CA..CAGCC.G..G...AG..GT...T.....CA.....
P2aBBGc-8(1,1) GTCATTGGCCACTTACCAGAACA....GATT.GG..TC.AG.TA..CA..CAGCC.G..G...AG..GT...T.....CA.....
P2aBBGc-6(2,1) NNNNNNNN
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
KC955130_BG8 CTATTTTGCAGAGAGAAAAGATGCAGAGTTGGGTAAGTCTCCTTCCCTAAAGCGAGGGAATTCAGGGTGTCCCCATGGCATCAGCCGTGGAATTAGTAGC
P2aBBGa(1,1) .....
P2aBBGb(4,1) .....
P2aBBGc-1(13,1st) .....
P2aBBGc-3(8,2nd) T.....G.....CAC...
P2aBBGc-2(13,1) ..G.....CAC...
P2aBBGc-4(8,T) ..G.....CAC...
P2aBBGc-5(2,2) T.....G.....CAC...
P2aBBGc-7(1,1) T.....G.....CAC...
P2aBBGc-8(1,1) T.....G.....CAC...
P2aBBGc-6(2,1)
P2aBBGc-14(1,1) ..G.....CAC...
P2aBBGc-10(1,1) ..G.....CAC...
P2aBBGc-15(1,1) ..G.....CAC...
P2aBBGc-13(1,1) ..G.....CAC...
P2aBBGc-11(1,1) ..G.....CAC...
P2aBBGc-9(1,1) ..G.....CAC...
P2aBBGc-16(20,V)

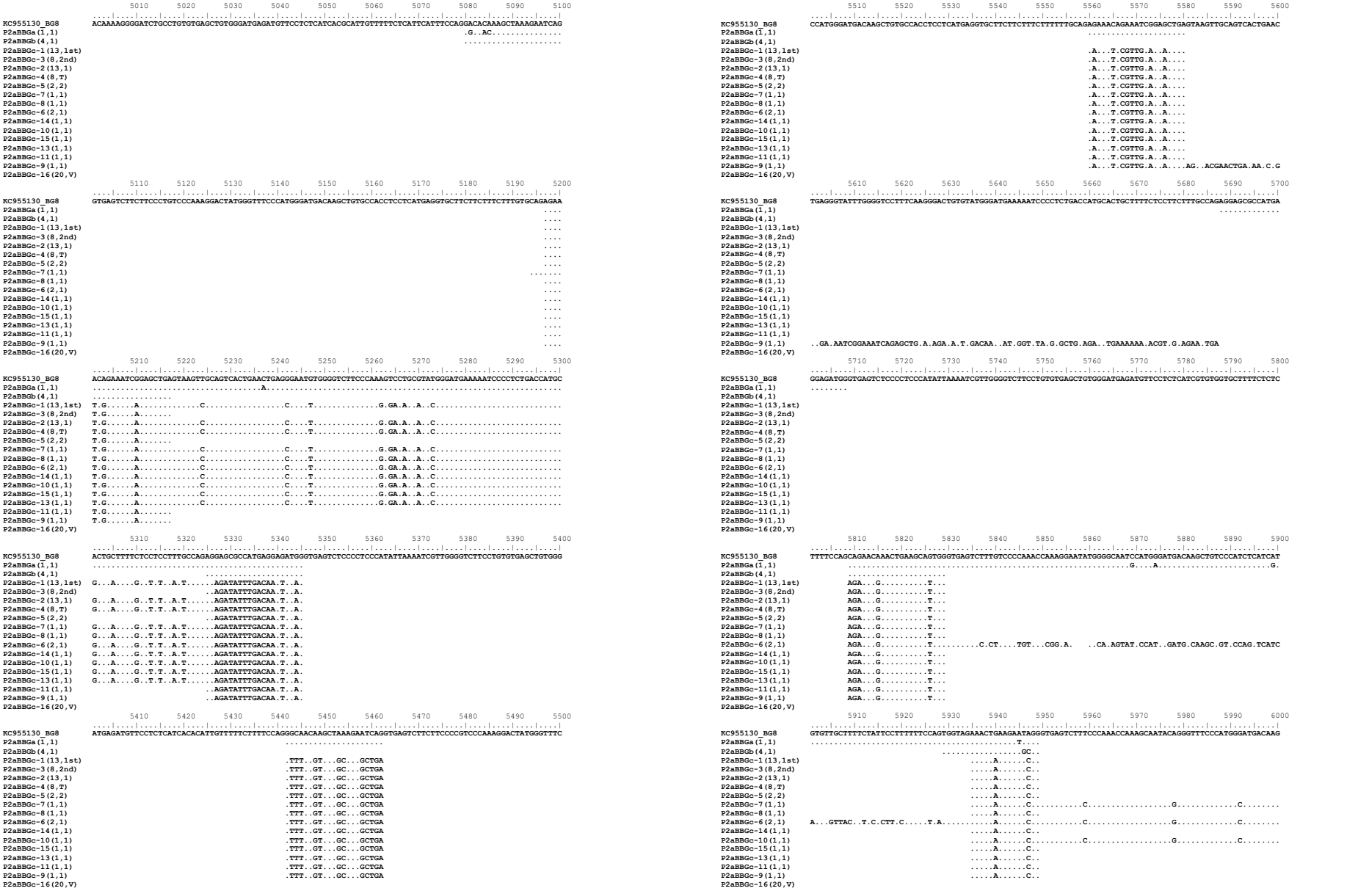

KC955130_BG8
P2aBBGa(1,1)
P2aBBGb(4,1)
P2aBBGc-1(13,1st)
P2aBBGc-3(8,2nd)
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2)
P2aBBGc-7(1,1)
P2aBBGc-8(1,1)
P2aBBGc-6(2,1)
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)

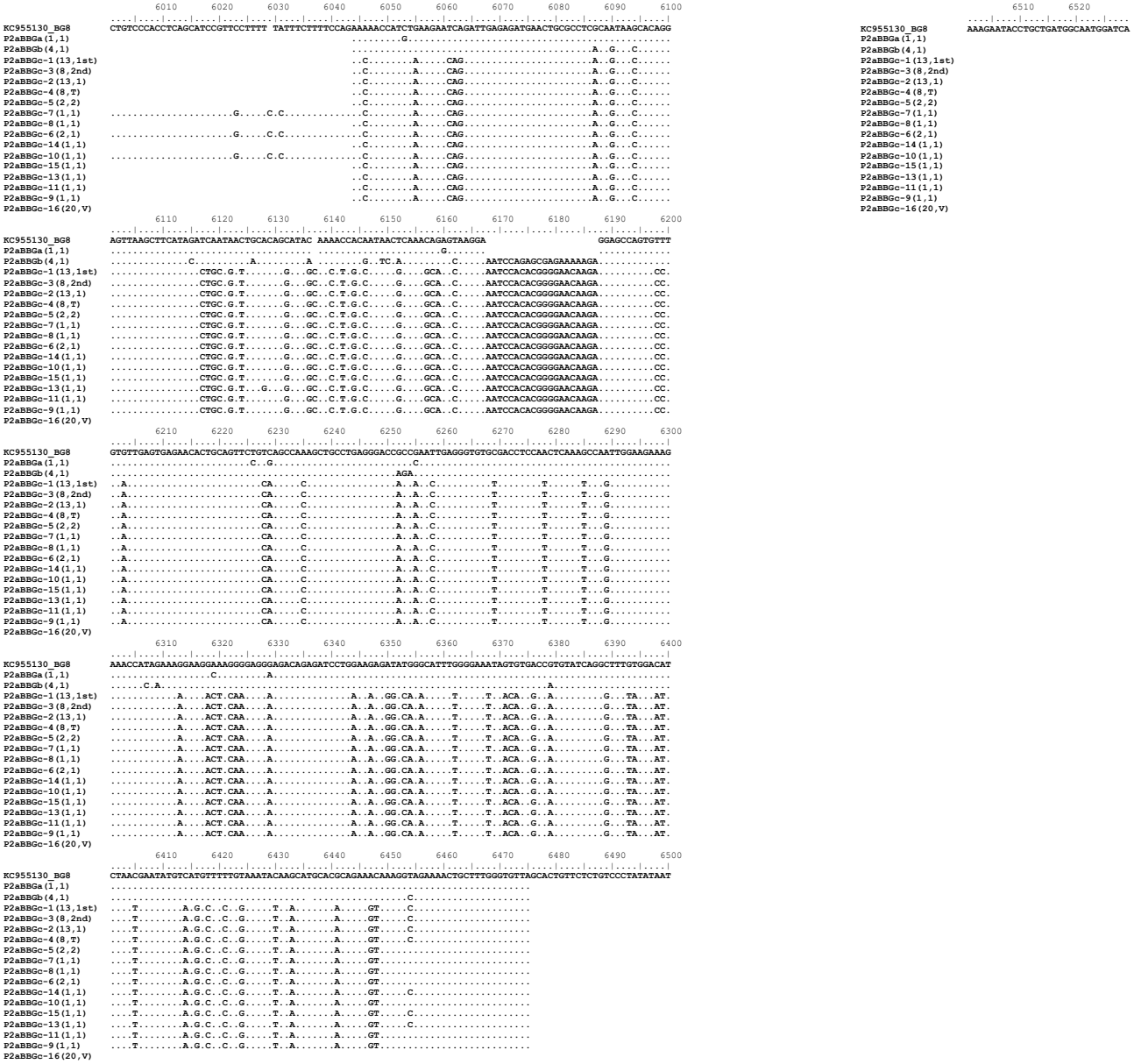

KC955130_BG8
P2aBBGa(1,1)
P2aBBGb(4,1)
P2aBBGc-1(13,1st)
P2aBBGc-3(8,2nd)
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2)
P2aBBGc-7(1,1)
P2aBBGc-8(1,1)
P2aBBGc-6(2,1)
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)


KC955130_BG8
P2aBBGa(1,1)
P2aBBGb(4,1)
P2aBBGc-1(13,1st)
P2aBBGc-3(8,2nd)
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2)
P2aBBGc-7(1,1)
P2aBBGc-8(1,1)
P2aBBGc-6(2,1)
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)

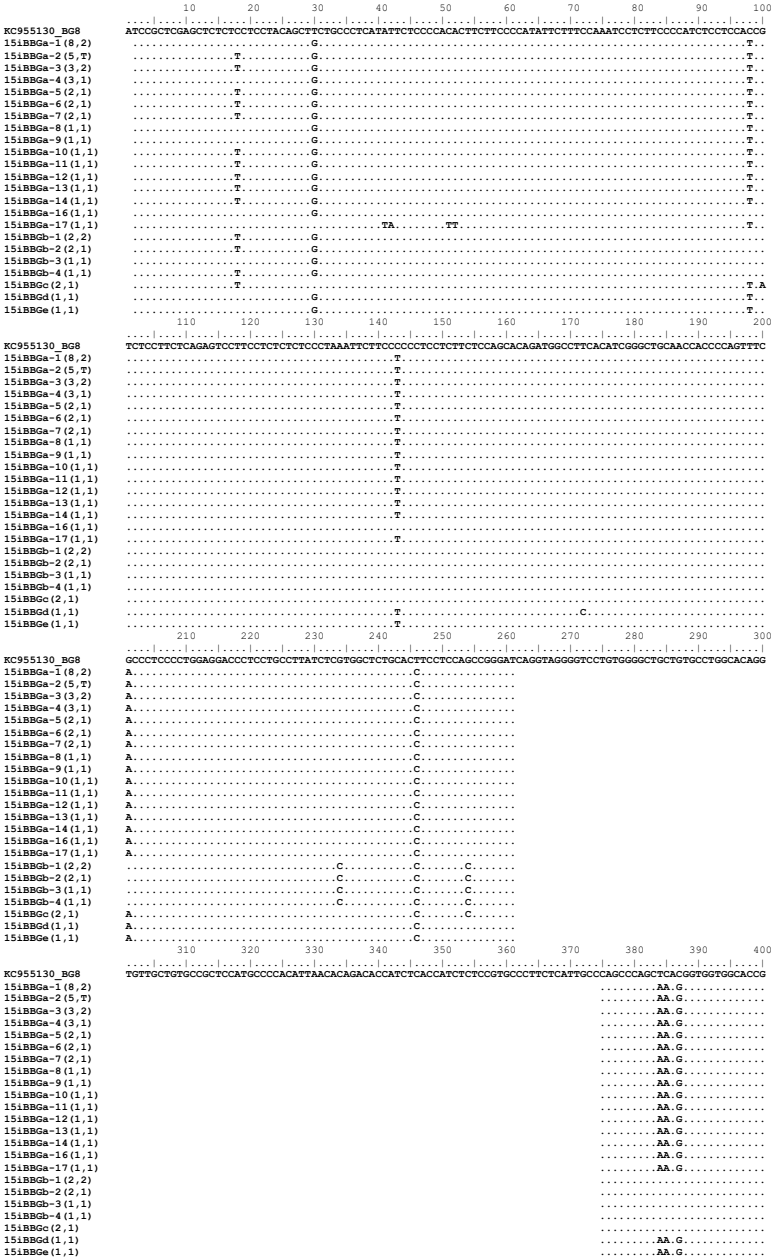

KC955130_BG8
P2aBBGa(1,1)
P2aBBGb(4,1)
P2aBBGc-1(13,1st)
P2aBBGc-3(8,2nd)
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2)
P2aBBGc-7(1,1)
P2aBBGc-8(1,1)
P2aBBGc-6(2,1)
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)

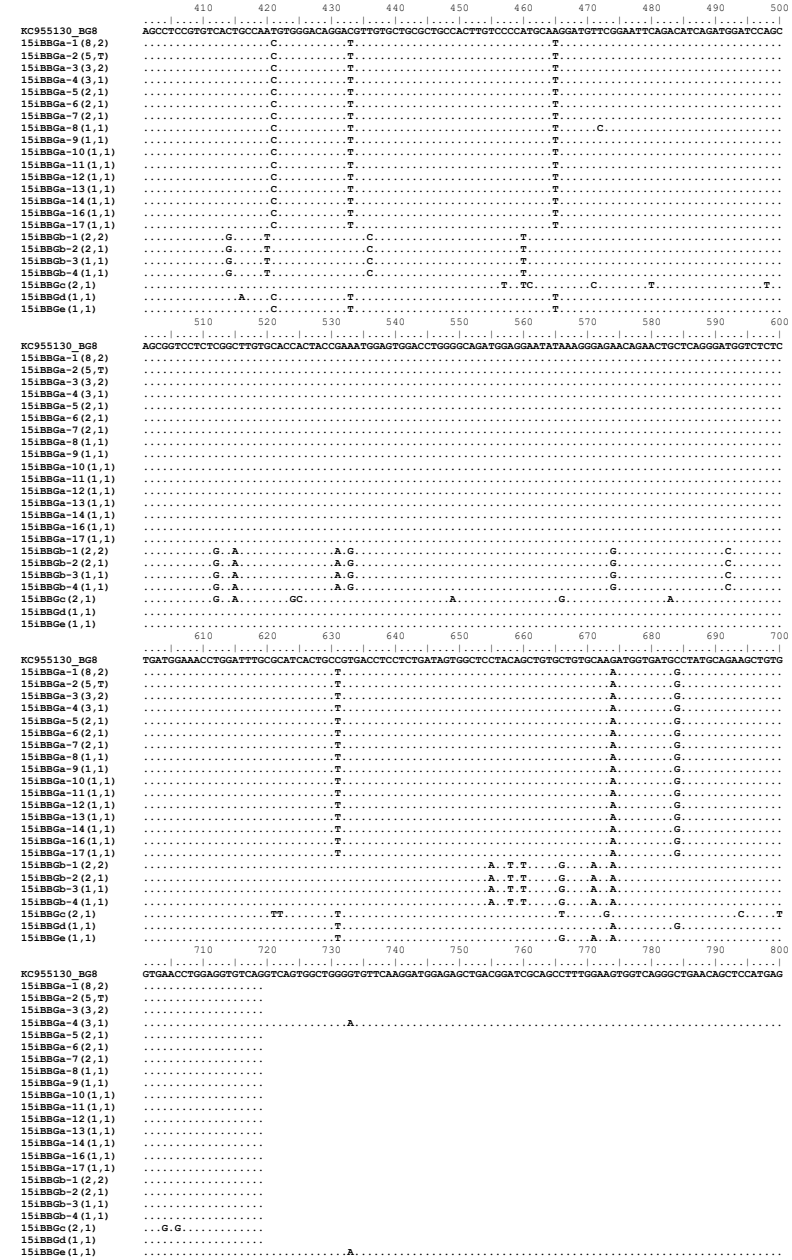

4010 4020 4030 4040 4050 4060 4070 4080 4090 4100
KC955130_BG8 GTCTTCCCAAAGTCCTGCATGTGGGATGAAAAATCCCCTCTGACCATGCACTGCTTTTCTCCTTCTATTCCAGAGAGACACTTTCAGAATATGTGTGAGT
P2aBBGa(1,1)
P2aBBGb(4,1)
P2aBBGc-1(13,1st) ...ACG.AA..G.AG.AG..G.....
P2aBBGc-3(8,2nd) ...ACG.AA..G.AG.AG..G
P2aBBGc-2(13,1) ...ACG.AA..G.AG.AG..G
P2aBBGc-4(8,T) ...ACG.AA..G.AG.AG..G
P2aBBGc-5(2,2) ...ACG.AA..G.AG.AG..G
P2aBBGc-7(1,1) ...ACG.AA..G.AG.AG..G
P2aBBGc-8(1,1) ...ACG.AA..G.AG.AG..G
P2aBBGc-6(2,1) ...ACG.AA..G.AG.AG..G
P2aBBGc-14(1,1) ...ACG.AA..G.AG.AG..G
P2aBBGc-10(1,1) ...ACG.AA..G.AG.AG..G
P2aBBGc-15(1,1) ...ACG.AA..G.AG.AG..G
P2aBBGc-13(1,1) ...ACG.AA..G.AG.AG..G
P2aBBGc-11(1,1) ...ACG.AA..G.AG.AG..G
P2aBBGc-9(1,1) ...ACG.AA..G.AG.AG..G
P2aBBGc-16(20,V)
4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
KC955130_BG8 CTCCCCACCCCTGATAAATAAAAACGTTGGGGTCTTGCTGTGTGAGCTGTGGGATGAGATGTTCCTCTCATCACACATTGTTTTTCTGATTTTCTTTTGC
P2aBBGa(1,1)
P2aBBGb(4,1)
P2aBBGc-1(13,1st) ..T.TTC..TGAACCG..GG..TTT.GG.TCT..CCAAG.GACA.CA.A....C...A....C.. ...TCAC..TG.AC.GA..T.CTC.......CA.
P2aBBGc-3(8,2nd)
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2)
P2aBBGc-7(1,1)
P2aBBGc-8(1,1)
P2aBBGc-6(2,1)
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)
4210 4220 4230 4240 4250 4260 4270 4280 4290 4300
KC955130_BG8 AGATTTAAGTGCTGGAAAACAGAGTAAGTCTCCCTCCCTGCACAGAAGGAACTTACGGTTTTCCCATGGGATCAGCCATGGGATCATCATCCGACTCTTC
P2aBBGa(1,1)
P2aBBGb(4,1) .....T..T.....
P2aBBGc-1(13,1st) ...GAA.CACCT.AA...GACTG
P2aBBGc-3(8,2nd) .GAA.CACCT.AA...GACTG
P2aBBGc-2(13,1) .GAA.CACCT.AA...GACTG
P2aBBGc-4(8,T) .GAA.CACCT.AA...GACTG
P2aBBGc-5(2,2) .GAA.CACCT.AA...GACTG
P2aBBGc-7(1,1) .GAA.CACCT.AA...GACTG
P2aBBGc-8(1,1) .GAA.CACCT.AA...GACTG
P2aBBGc-6(2,1) .GAA.CACCT.AA...GACTG
P2aBBGc-14(1,1) .GAA.CACCT.AA...GACTG
P2aBBGc-10(1,1) .GAA.CACCT.AA...GACTG
P2aBBGc-15(1,1) .GAA.CACCT.AA...GACTG
P2aBBGc-13(1,1) .GAA.CACCT.AA...GACTG
P2aBBGc-11(1,1) .GAA.CACCT.AA...GACTG
P2aBBGc-9(1,1) .GAA..CAC..TA....G.CT
P2aBBGc-16(20,V)
4310 4320 4330 4340 4350 4360 4370 4380 4390 4400
KC955130_BG8 TCATCATGAATTTCGTCTTTCTTTCTTTTGCAGAGAAAATGGTTACAAAACTGGGTGAGTCCAACCTCCCAAACTAAATTAAAAACAGTCAGACTTTGTG
P2aBBGa(1,1)
P2aBBGb(4,1)
P2aBBGc-1(13,1st) GT.G.CGT.C.C.T..T...A
P2aBBGc-3(8,2nd) GT.G.CGT.C.C.T..T...A
P2aBBGc-2(13,1) GT.G.CGT.C.C.T..T...A
P2aBBGc-4(8,T) GT.G.CGT.C.C.T..T...A
P2aBBGc-5(2,2) GT.G.CGT.C.C.T..T...A
P2aBBGc-7(1,1) GT.G.CGT.C.C.T..T...A
P2aBBGc-8(1,1) GT.G.CGT.C.C.T..T...A
P2aBBGc-6(2,1) GT.G.CGT.C.C.T..T...A
P2aBBGc-14(1,1) GT.G.CGT.C.C.T..T...A
P2aBBGc-10(1,1) GT.G.CGT.C.C.T..T...A
P2aBBGc-15(1,1) GT.G.CGT.C.C.T..T...A
P2aBBGc-13(1,1) GT.G.CGT.C.C.T..T...A
P2aBBGc-11(1,1) GT.G.CGT.C.C.T..T...A
P2aBBGc-9(1,1) G.CGT.C.C.T..T...A
P2aBBGc-16(20,V)
4410 4420 4430 4440 4450 4460 4470 4480 4490 4500
KC955130_BG8 AGCTGTGGGATGAGACGTTCCTCTCATCATGTGCTGCTTTCCTTTTACTTTTCCAGAGGAACACTGTGAATGGATGGGTGAGTCTCCCCTCCCAAATTAA
P2aBBGa(1,1)
P2aBBGb(4,1) ...C.....CA...
P2aBBGc-1(13,1st) ..C.....ATG.C.GAAC.....GCT.....
P2aBBGc-3(8,2nd) ..C.....ATG.C.GAAC...
P2aBBGc-2(13,1) ..C.....ATG.C.GAAC...
P2aBBGc-4(8,T) ..C.....ATG.C.GAAC...
P2aBBGc-5(2,2) ..C.....ATG.C.GAAC...
P2aBBGc-7(1,1)
P2aBBGc-8(1,1) ..C.....ATG.C.GAAC...
P2aBBGc-6(2,1) ..C.....ATG.C.GAAC...
P2aBBGc-14(1,1) ..C.....ATG.C.GAAC...
P2aBBGc-10(1,1) ..C.....ATG.C.GAAC...
P2aBBGc-15(1,1) ..C.....ATG.C.GAAC...
P2aBBGc-13(1,1) ..C.....ATG.C.GAAC...
P2aBBGc-11(1,1) ..C.....ATG.C.GAAC...
P2aBBGc-9(1,1) ..C.....ATG.C.GAAC...
P2aBBGc-16(20,V)
4510 4520 4530 4540 4550 4560 4570 4580 4590 4600
KC955130_BG8 AAATGTTGGGGTCTTCCTGTGAGAGCTGTGGGATGAGCTGTTCCTCTCATCGTGCACTGTTTCTGCTTTTCCTTTGCAGTGAGAAGGAATGTAAAGTTGG
P2aBBGa(1,1)
P2aBBGb(4,1) .....C..G....
P2aBBGc-1(13,1st) C..CAAA...C...G.....TC....C.T.......AG.C.........A....T..C..T.CTC..C.A...C.....GATCAAGC..A.GCAG...
P2aBBGc-3(8,2nd) ..GATCAAGC..A.GCAG...
P2aBBGc-2(13,1) ..GATCAAGC..A.GCAG...
P2aBBGc-4(8,T) ..GATCAAGC..A.GCAG...
P2aBBGc-5(2,2) ..GATCAAGC..A.GCAG...
P2aBBGc-7(1,1)
P2aBBGc-8(1,1) ..GATCAAGC..A.GCAG...
P2aBBGc-6(2,1) ..GATCAAGC..A.GCAG...
P2aBBGc-14(1,1) ..GATCAAGC..A.GCAG...
P2aBBGc-10(1,1) ..GATCAAGC..A.GCAG...
P2aBBGc-15(1,1) ..GATCAAGC..A.GCAG...
P2aBBGc-13(1,1) ..GATCAAGC..A.GCAG...
P2aBBGc-11(1,1) ..GATCAAGC..A.GCAG...
P2aBBGc-9(1,1) ..GATCAAGC..A.GCAG...
P2aBBGc-16(20,V)
4610 4620 4630 4640 4650 4660 4670 4680 4690 4700
KC955130_BG8 GTGAGTCTTCTTCCCCAACCAAAGAGATTCGGAGTCTTCCATGGGATCAGCCATGGGATGATAACATGAACCTCATCACGTGTTTCTTATTTGTTCCTTT
P2aBBGa(1,1)
P2aBBGb(4,1)
P2aBBGc-1(13,1st) .....AGC..AG..A.TT...GT..T.CAT..GATGA.A.GAT.TCCCACCTCAAC..GCAG.GCTT.T.G.T.CT...T..TT...
P2aBBGc-3(8,2nd)
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2)
P2aBBGc-7(1,1)
P2aBBGc-8(1,1)
P2aBBGc-6(2,1)
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)
4710 4720 4730 4740 4750 4760 4770 4780 4790 4800
KC955130_BG8 TGCAGAGGCAGCAGCTGTAAAAGTGGGTGAGTCTCCCCTCCCAAATTAAAAATGTTGGCGTCATCCTGTGAGAGCTGTGGGATGAGCTGTTCCTCTCATC
P2aBBGa(1,1) ..AT.C.....
P2aBBGb(4,1) .....
P2aBBGc-1(13,1st) .....
P2aBBGc-3(8,2nd)
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2)
P2aBBGc-7(1,1)
P2aBBGc-8(1,1)
P2aBBGc-6(2,1)
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)
4810 4820 4830 4840 4850 4860 4870 4880 4890 4900
KC955130_BG8 GTGCACTGTTTCTGCTTTTCCTTTGCAGTGAGAAGGAATGTAAAGTTGGGTGAGTCTTCTTCCCCAACCAAAGAGATGTGGGGTCTTCCATGGGATCAGC
P2aBBGa(1,1) .....
P2aBBGb(4,1)
P2aBBGc-1(13,1st)
P2aBBGc-3(8,2nd)
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2)
P2aBBGc-7(1,1)
P2aBBGc-8(1,1)
P2aBBGc-6(2,1)
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)
4910 4920 4930 4940 4950 4960 4970 4980 4990 5000
KC955130_BG8 CATGGGATGATAAGCTGAACCTTATCACGTGTTTCTTATTTGTTCCTTTTGCAGAGGCAGCAGCTGTAAAAGTGGGTGAGTCCTCCCTCCCAAATCAAAT
P2aBBGa(1,1) ..AT.C.....
P2aBBGb(4,1)
P2aBBGc-1(13,1st)
P2aBBGc-3(8,2nd)
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2)
P2aBBGc-7(1,1)
P2aBBGc-8(1,1)
P2aBBGc-6(2,1)
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)

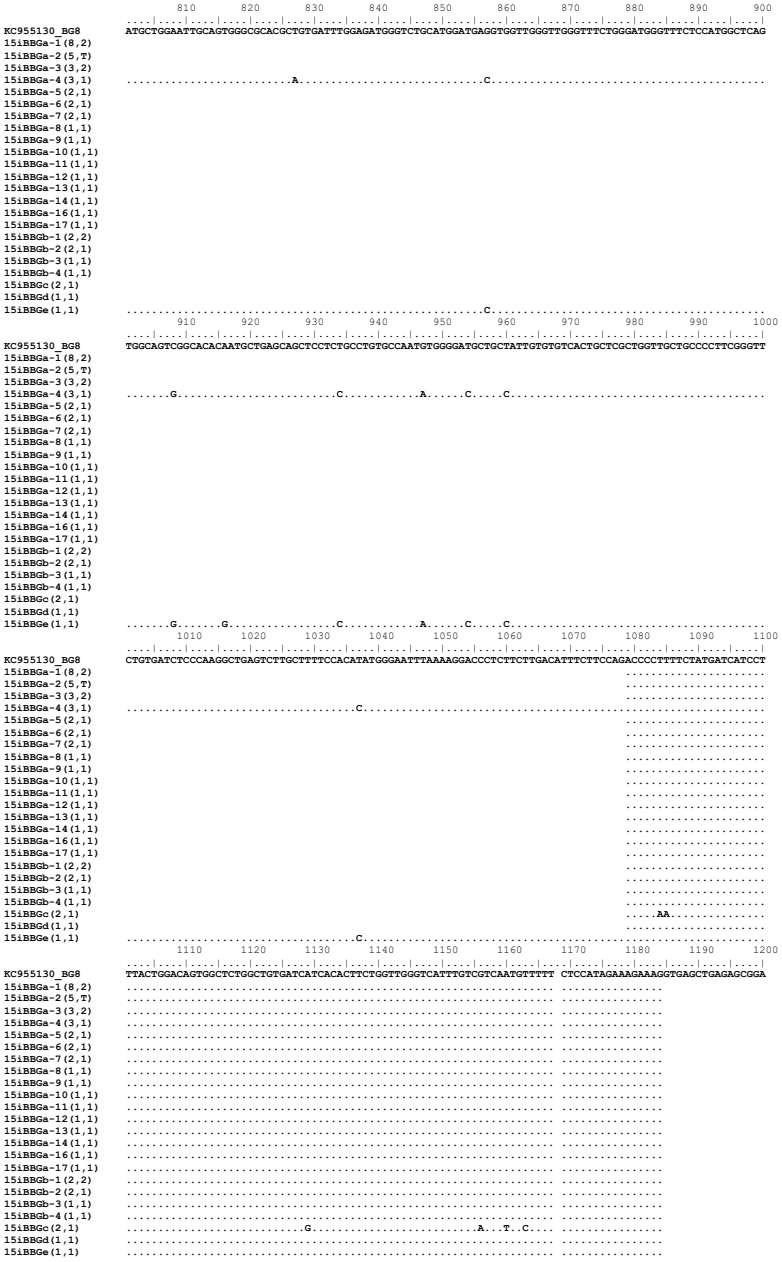

KC955130_BG8
P2aBBGa(1,1)
P2aBBGb(4,1)
P2aBBGc-1(13,1st)
P2aBBGc-3(8,2nd)
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2)
P2aBBGc-7(1,1)
P2aBBGc-8(1,1)
P2aBBGc-6(2,1)
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)

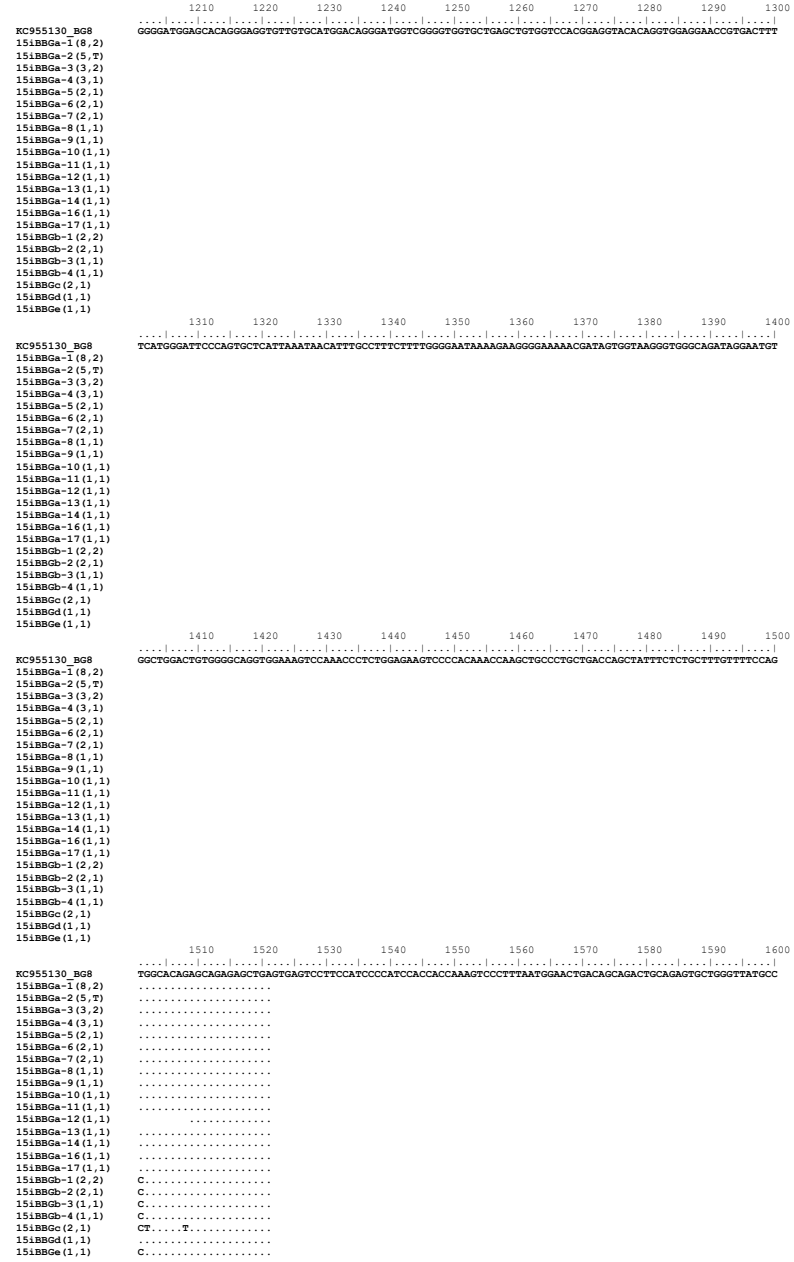

6010 6020 6030 6040 6050 6060 6070 6080 6090 6100
KC955130_BG8 CTGTCCCACCTCAGCATCCGTTCCTTTT TATTTCTTTTCCAGAAAAACCATCTGAAGAATCAGATTGAGAGATGAACTGCGCCTCGCAATAAGCACAGG
P2aBBGa(1,1)
P2aBBGb(4,1)
P2aBBGc-1(13,1st)
P2aBBGc-3(8,2nd)
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2)
P2aBBGc-7(1,1)
P2aBBGc-8(1,1)
P2aBBGc-6(2,1)
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)
6110 6120 6130 6140 6150 6160 6170 6180 6190 6200
KC955130_BG8 AGTTAAGCTTCATAGATCAATAACTGCACAGCATAC AAAACCACAATAACTCAAACAGAGTAAGGA GGAGCCAGTGTTT
P2aBBGa(1,1)
P2aBBGb(4,1)
P2aBBGc-1(13,1st)
P2aBBGc-3(8,2nd)
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2)
P2aBBGc-7(1,1)
P2aBBGc-8(1,1)
P2aBBGc-6(2,1)
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)
6210 6220 6230 6240 6250 6260 6270 6280 6290 6300
KC955130_BG8 GTGTTGAGTGAGAACACTGCAGTTCTGTCAGCCAAAGCTGCCTGAGGGACCGCCGAATTGAGGGTGTGCGACCTCCAACTCAAAGCCAATTGGAAGAAAG
P2aBBGa(1,1)
P2aBBGb(4,1)
P2aBBGc-1(13,1st)
P2aBBGc-3(8,2nd)
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2)
P2aBBGc-7(1,1)
P2aBBGc-8(1,1)
P2aBBGc-6(2,1)
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)
6310 6320 6330 6340 6350 6360 6370 6380 6390 6400
KC955130_BG8 AAACCATAGAAAGGAAGGAAAGGGGAGGGAGACAGAGATCCTGGAAGAGATATGGGCATTTGGGGAAATAGTGTGACCGTGTATCAGGCTTTGTGGACAT
P2aBBGa(1,1)
P2aBBGb(4,1)
P2aBBGc-1(13,1st)
P2aBBGc-3(8,2nd)
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2)
P2aBBGc-7(1,1)
P2aBBGc-8(1,1)
P2aBBGc-6(2,1)
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)
6410 6420 6430 6440 6450 6460 6470 6480 6490 6500
KC955130_BG8 CTAACGAATATGTCATGTTTTTGTAAATACAAGCATGCACGCAGAAACAAAGGTAGAAAACTGCTTTGGGTGTTAGCACTGTTCTCTGTCCCTATATAAT
P2aBBGa(1,1)
P2aBBGb(4,1)
P2aBBGc-1(13,1st)
P2aBBGc-3(8,2nd)
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2)
P2aBBGc-7(1,1)
P2aBBGc-8(1,1)
P2aBBGc-6(2,1)
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)
6510 6520
KC955130_BG8 AAAGAATACCTGCTGATGGCAATGGATCA
P2aBBGa(1,1)
P2aBBGb(4,1)
P2aBBGc-1(13,1st)
P2aBBGc-3(8,2nd)
P2aBBGc-2(13,1)
P2aBBGc-4(8,T)
P2aBBGc-5(2,2)
P2aBBGc-7(1,1)
P2aBBGc-8(1,1)
P2aBBGc-6(2,1)
P2aBBGc-14(1,1)
P2aBBGc-10(1,1)
P2aBBGc-15(1,1)
P2aBBGc-13(1,1)
P2aBBGc-11(1,1)
P2aBBGc-9(1,1)
P2aBBGc-16(20,V)

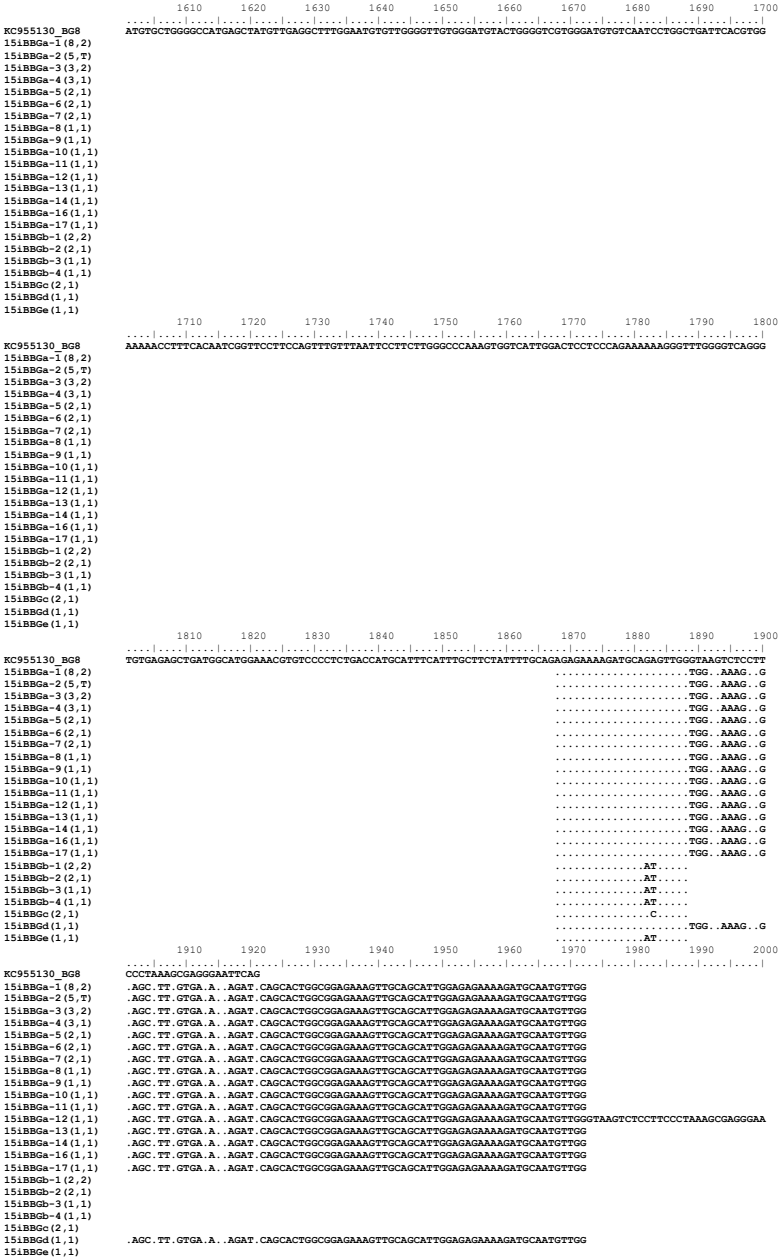
(J3)

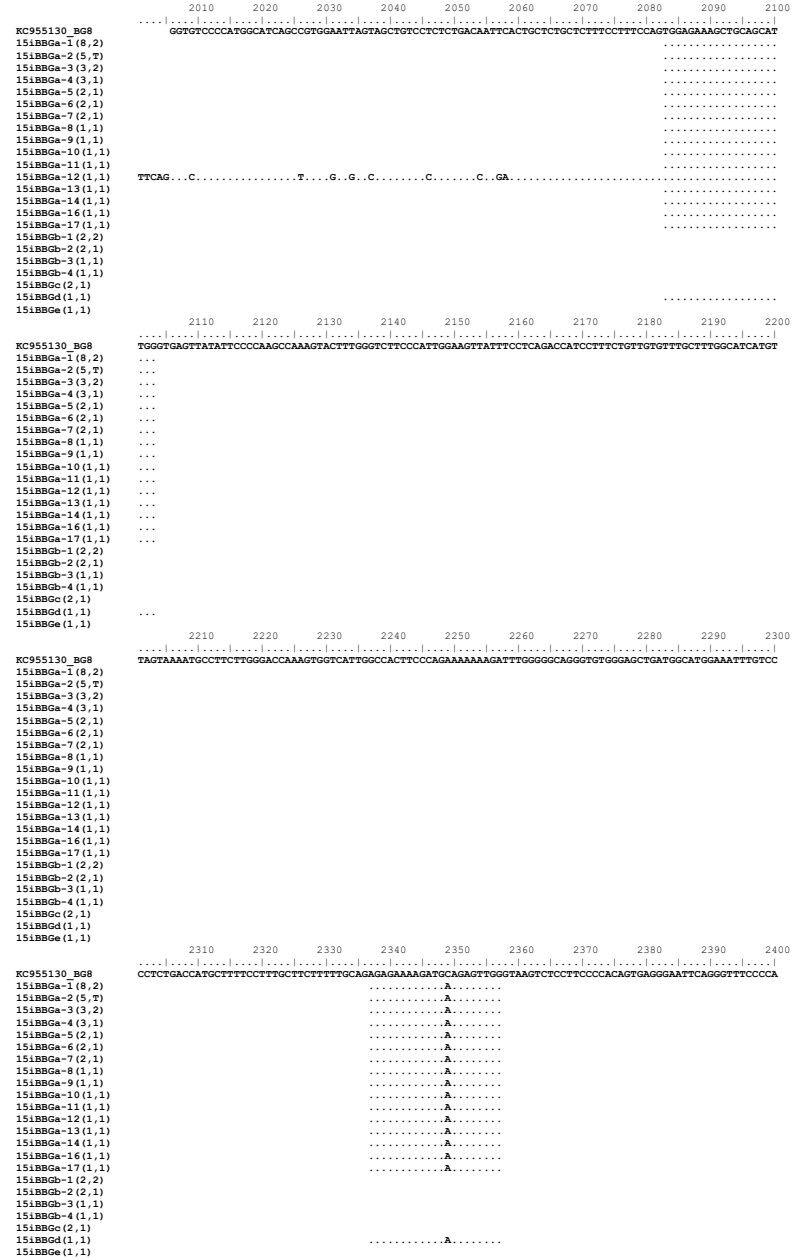

KC955130_BG8
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1)
15iBBGe(1,1)

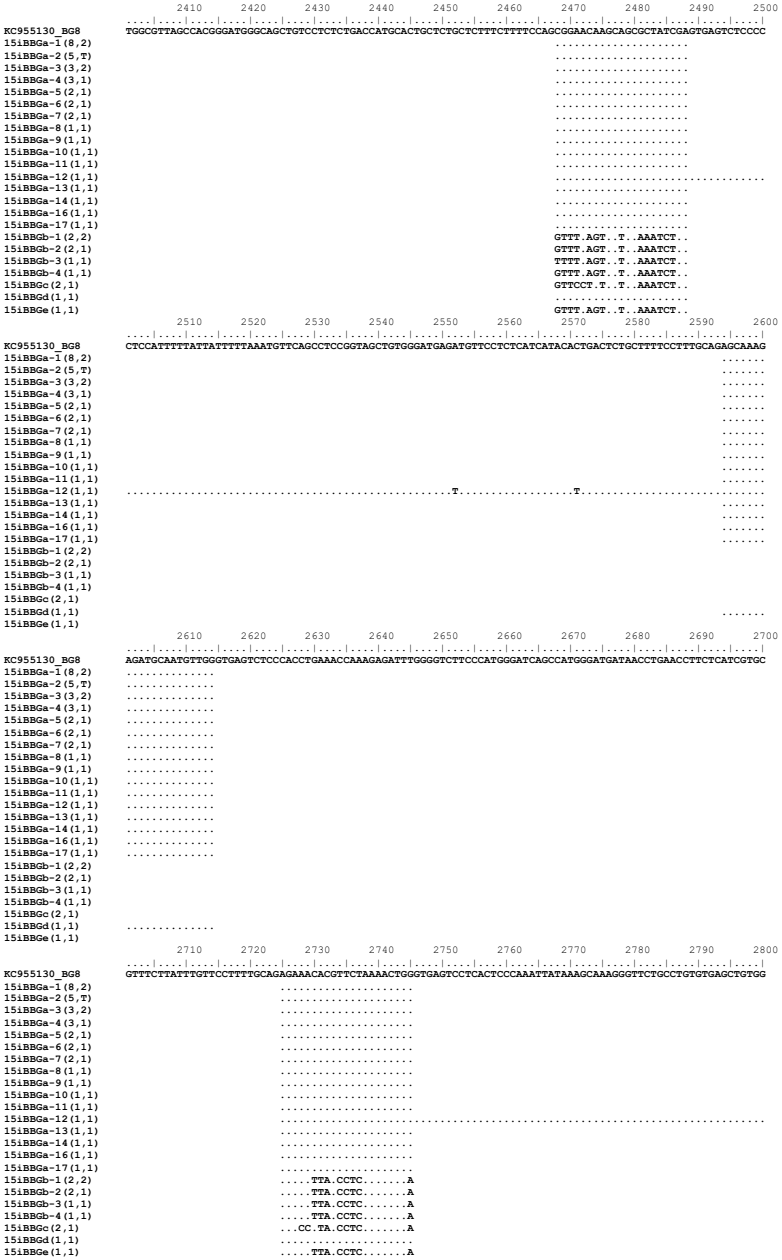

KC955130_BG8
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1)
15iBBGe(1,1)

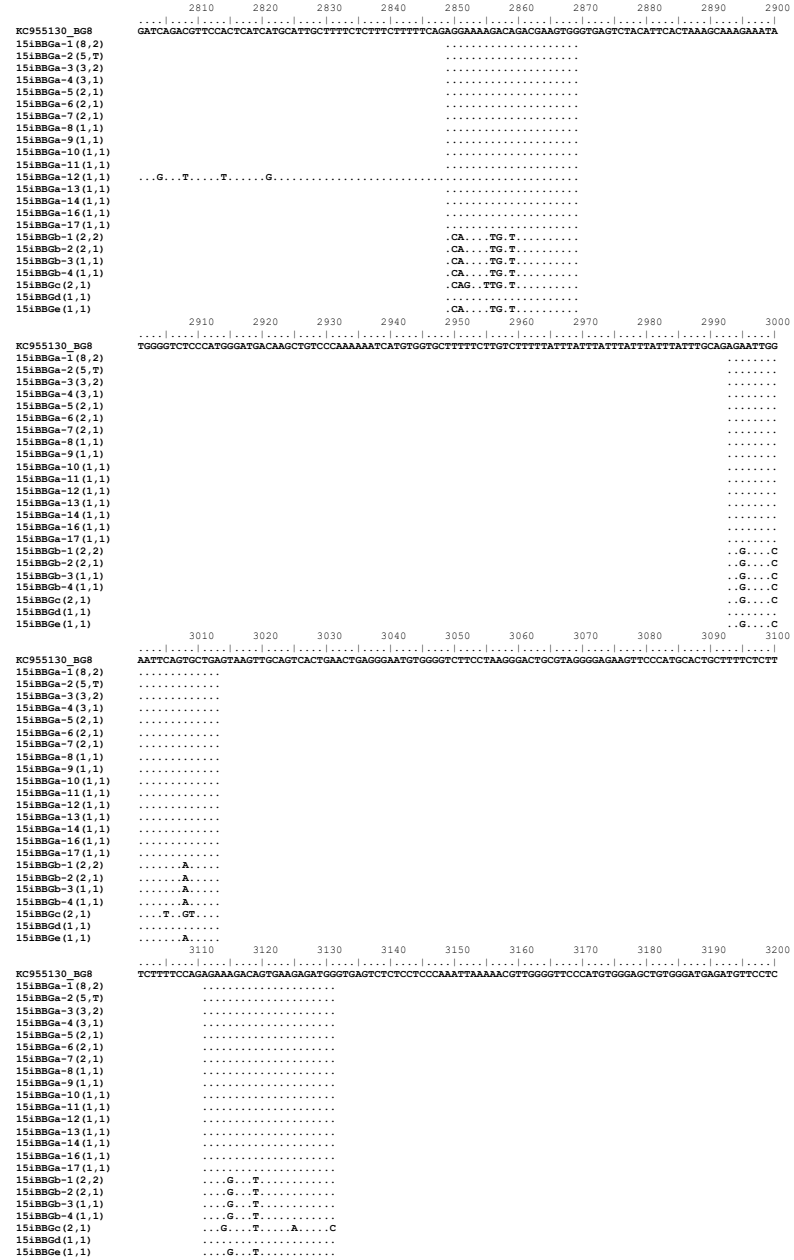

KC955130_BG8
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1)
15iBBGe(1,1)

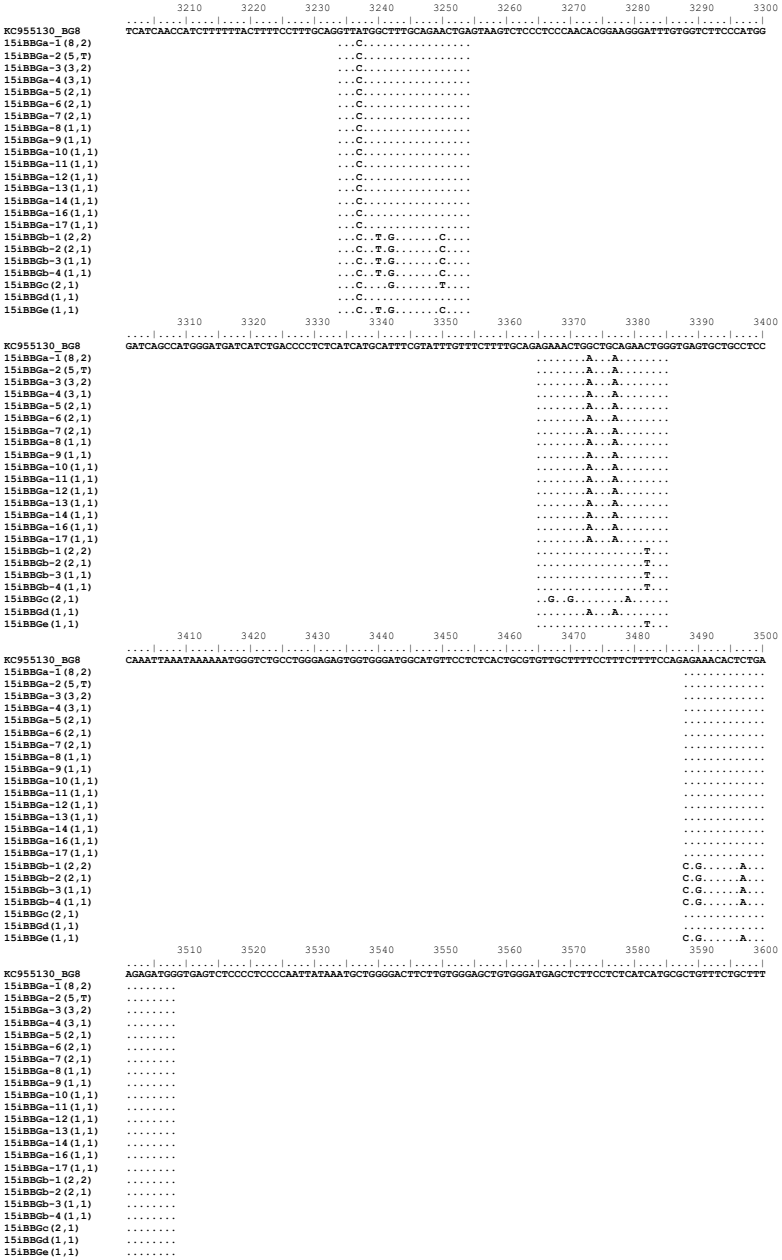

KC955130_BG8
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1)
15iBBGe(1,1)

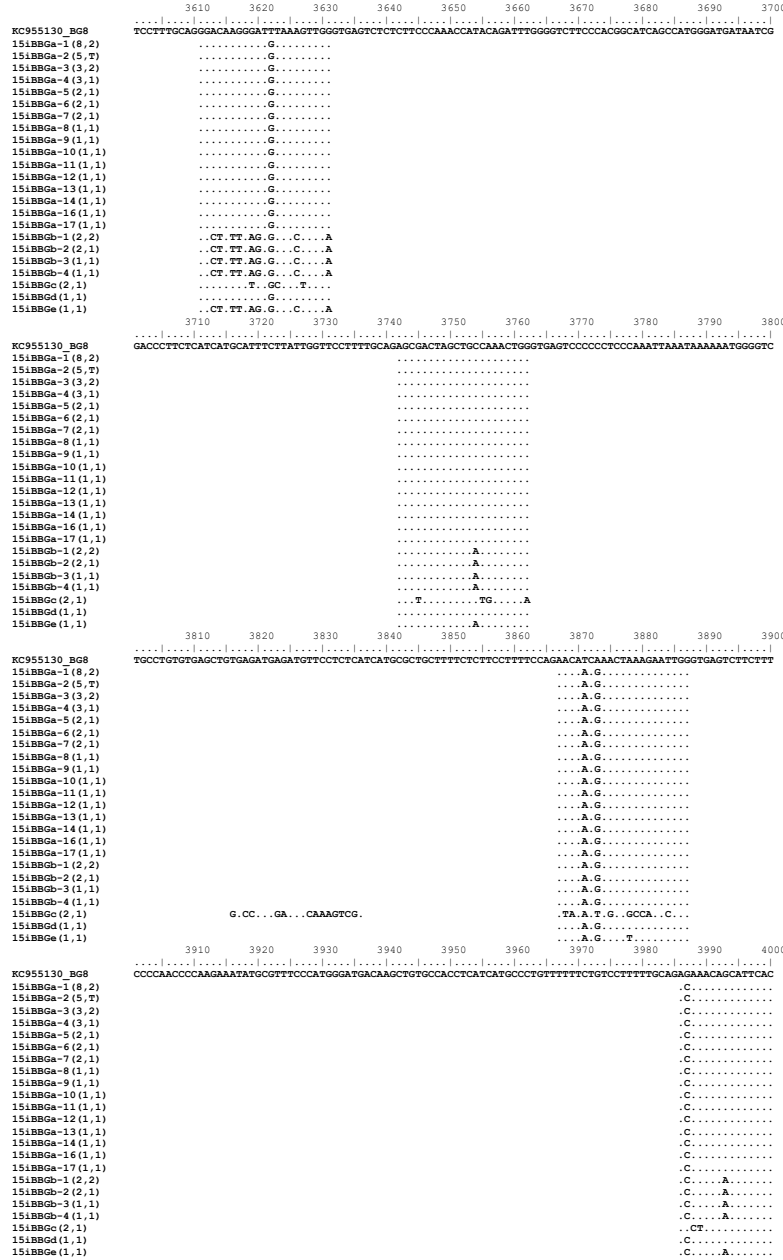

1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
KC955130_BG8 ATGTGCTGGGGCCATGAGCTATGTTGAGGCTTTGGAATGTGTTGGGGTTGTGGGATGTACTGGGGTCGTGGGATGTGTCAATCCTGGCTGATTCACGTGG
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
KC955130_BG8 AAAAACCTTTCACAATCGGTTCCTTCCAGTTTGTTTAATTCCTTCTTGGGCCCAAAGTGGTCATTGGACTCCTCCCAGAAAAAAGGGTTTGGGGTCAGGG
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
KC955130_BG8 TGTGAGAGCTGATGGCATGGAAACGTGTCCCCTCTGACCATGCATTTCATTTGCTTCTATTTTGCAGAGAGAAAAAGATGCAGAGTTGGGTAAGTCTCCTT
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
KC955130_BG8 CCCTAAAGCGAGGGAATTCAG
15iBBGa-1(8,2) .AGC.TT.GTGA.A..AGAT.CAGCACTGGCGGAGAAAGTTGCAGCATTGGAGAGAAAAGATGCAATGTTGG
15iBBGa-2(5,T) .AGC.TT.GTGA.A..AGAT.CAGCACTGGCGGAGAAAGTTGCAGCATTGGAGAGAAAAGATGCAATGTTGG
15iBBGa-3(3,2) .AGC.TT.GTGA.A..AGAT.CAGCACTGGCGGAGAAAGTTGCAGCATTGGAGAGAAAAGATGCAATGTTGG
15iBBGa-4(3,1) .AGC.TT.GTGA.A..AGAT.CAGCACTGGCGGAGAAAGTTGCAGCATTGGAGAGAAAAGATGCAATGTTGG
15iBBGa-5(2,1) .AGC.TT.GTGA.A..AGAT.CAGCACTGGCGGAGAAAGTTGCAGCATTGGAGAGAAAAGATGCAATGTTGG
15iBBGa-6(2,1) .AGC.TT.GTGA.A..AGAT.CAGCACTGGCGGAGAAAGTTGCAGCATTGGAGAGAAAAGATGCAATGTTGG
15iBBGa-7(2,1) .AGC.TT.GTGA.A..AGAT.CAGCACTGGCGGAGAAAGTTGCAGCATTGGAGAGAAAAGATGCAATGTTGG
15iBBGa-8(1,1) .AGC.TT.GTGA.A..AGAT.CAGCACTGGCGGAGAAAGTTGCAGCATTGGAGAGAAAAGATGCAATGTTGG
15iBBGa-9(1,1) .AGC.TT.GTGA.A..AGAT.CAGCACTGGCGGAGAAAGTTGCAGCATTGGAGAGAAAAGATGCAATGTTGG
15iBBGa-10(1,1) .AGC.TT.GTGA.A..AGAT.CAGCACTGGCGGAGAAAGTTGCAGCATTGGAGAGAAAAGATGCAATGTTGG
15iBBGa-11(1,1) .AGC.TT.GTGA.A..AGAT.CAGCACTGGCGGAGAAAGTTGCAGCATTGGAGAGAAAAGATGCAATGTTGG
15iBBGa-12(1,1) .AGC.TT.GTGA.A..AGAT.CAGCACTGGCGGAGAAAGTTGCAGCATTGGAGAGAAAAGATGCAATGTTGGGTAAGTCTCCTTCCCTAAAGCGAGGGAA
15iBBGa-13(1,1) .AGC.TT.GTGA.A..AGAT.CAGCACTGGCGGAGAAAGTTGCAGCATTGGAGAGAAAAGATGCAATGTTGG
15iBBGa-14(1,1) .AGC.TT.GTGA.A..AGAT.CAGCACTGGCGGAGAAAGTTGCAGCATTGGAGAGAAAAGATGCAATGTTGG
15iBBGa-16(1,1) .AGC.TT.GTGA.A..AGAT.CAGCACTGGCGGAGAAAGTTGCAGCATTGGAGAGAAAAGATGCAATGTTGG
15iBBGa-17(1,1) .AGC.TT.GTGA.A..AGAT.CAGCACTGGCGGAGAAAGTTGCAGCATTGGAGAGAAAAGATGCAATGTTGG
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1) .AGC.TT.GTGA.A..AGAT.CAGCACTGGCGGAGAAAGTTGCAGCATTGGAGAGAAAAGATGCAATGTTGG
15iBBGe(1,1)

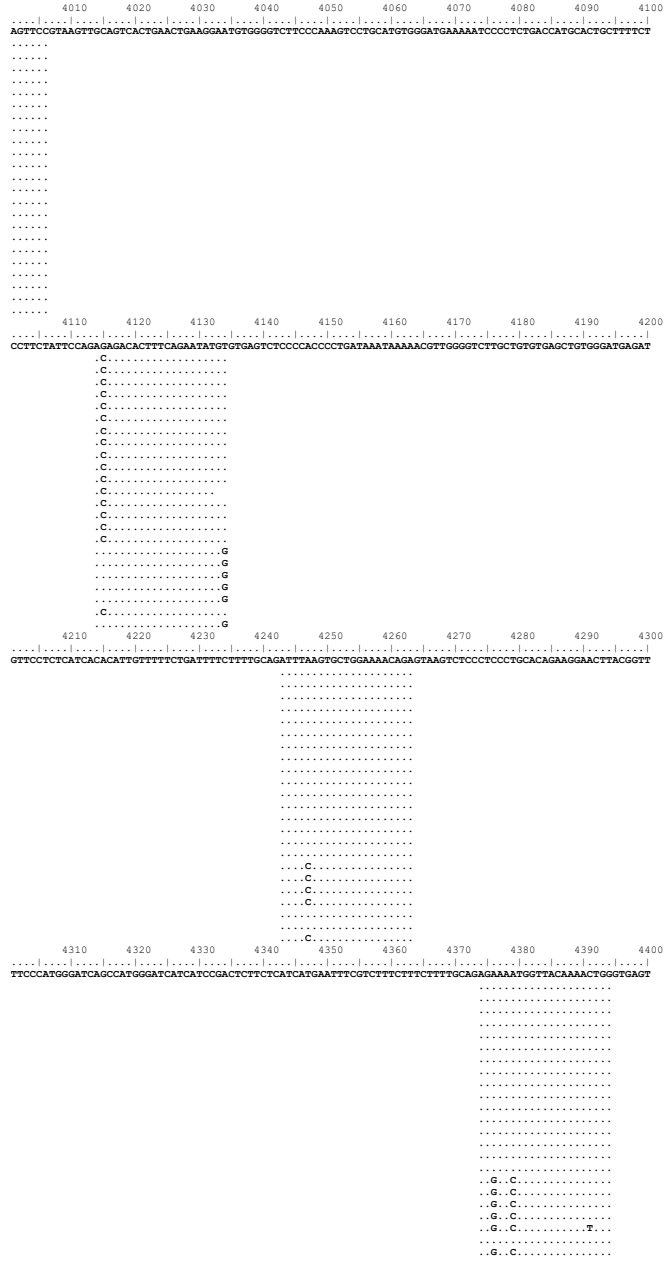

2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
KC955130_BG8 GGTGTCCCCATGGCATCAGCCGTGGAATTAGTAGCTGTCCTCTCTGACAATTCACTGCTCTGCTCTTTCCTTTCCAGTGGAGAAAGCTGCAGCAT
15iBBGa-12(1,1) TTCAG...C.............T....G..G..C.......C.......C..GA
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
KC955130_BG8 TGGGTGAGTTATATTCCCCAAGCCAAAGTACTTTGGGTCTTCCCATTGGAAGTTATTTCCTCAGACCATCCTTTCTGTTGTGTTTGCTTTGGCATCATGT
15iBBGa-1(8,2) ...
15iBBGa-2(5,T) ...
15iBBGa-3(3,2) ...
15iBBGa-4(3,1) ...
15iBBGa-5(2,1) ...
15iBBGa-6(2,1) ...
15iBBGa-7(2,1) ...
15iBBGa-8(1,1) ...
15iBBGa-9(1,1) ...
15iBBGa-10(1,1) ...
15iBBGa-11(1,1) ...
15iBBGa-12(1,1) ...
15iBBGa-13(1,1) ...
15iBBGa-14(1,1) ...
15iBBGa-16(1,1) ...
15iBBGa-17(1,1) ...
15iBBGd(1,1) ...
2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
KC955130_BG8 TAGTAAAATGCCTTCTTGGGACCAAAGTGGTCATTGGCCACTTCCCAGAAAAAAAGATTTGGGGGCAGGGTGTGGGAGCTGATGGCATGGAAATTTGTCC
2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
KC955130_BG8 CCTCTGACCATGCTTTTCCTTTGCTTCTTTTTGCAGAGAGAAAAGATGCAGAGTTGGGTAAGTCTCCTTCCCCACAGTGAGGGAATTCAGGGTTTCCCCA

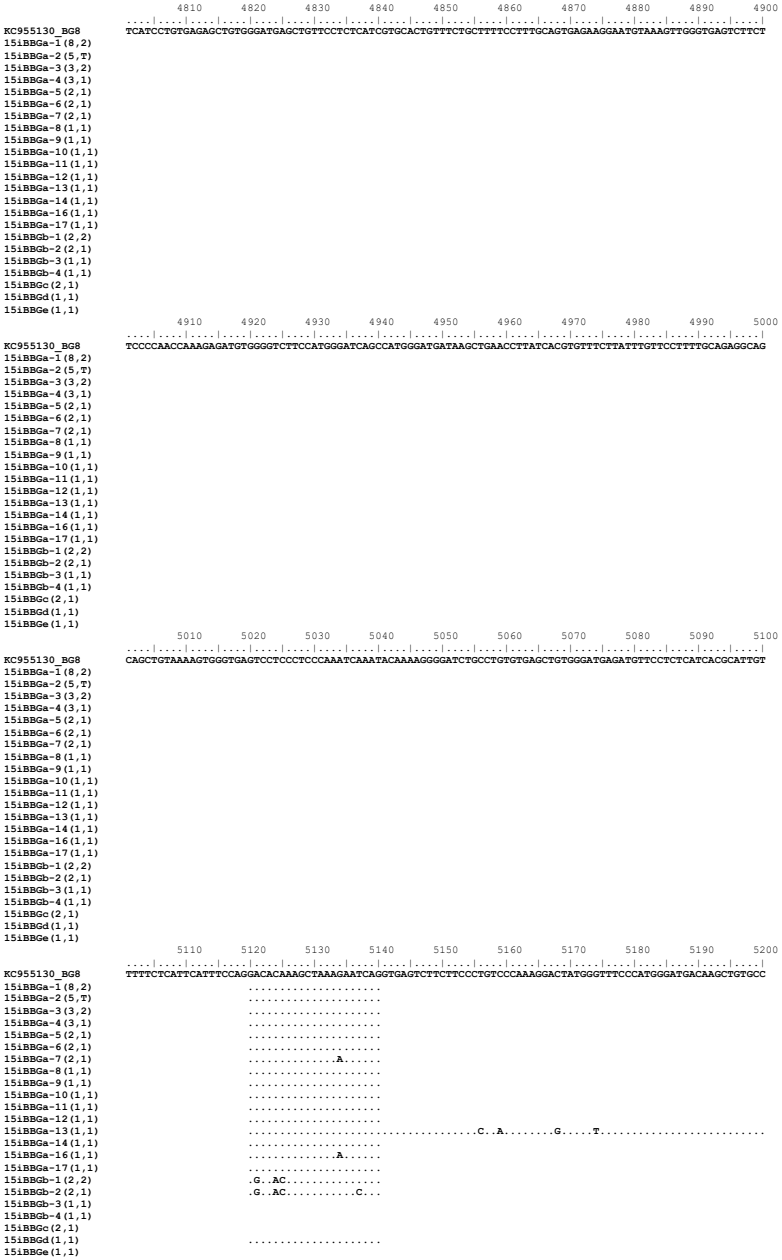

2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
KC955130_BG8 TGGCGTTAGCCACGGGATGGGCAGCTGTCCTCTCTGACCATGCACTGCTCTGCTCTTTCTTTTCCAGCGGAACAAGCAGCGCTATCGAGTGAGTCTCCCC
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2) GTTT.AGT..T..AAATCT..
15iBBGb-2(2,1) GTTT.AGT..T..AAATCT..
15iBBGb-3(1,1) TTTT.AGT..T..AAATCT..
15iBBGb-4(1,1) GTTT.AGT..T..AAATCT..
15iBBGc(2,1) GTTCCT.T..T..AAATCT..
15iBBGd(1,1)
15iBBGe(1,1) GTTT.AGT..T..AAATCT..
2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
KC955130_BG8 CTCCATTTTTATTATTTTTAAATGTTCAGCCTCCGGTAGCTGTGGGATGAGATGTTCCTCTCATCATACACTGACTCTGCTTTTCCTTTGCAGAGCAAAG
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1) .....T.....T.....
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1)
15iBBGe(1,1)
2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
KC955130_BG8 AGATGCAATGTTGGGTGAGTCTCCCACCTGAAACCAAAGAGATTTGGGGTCTTCCCATGGGATCAGCCATGGGATGATAACCTGAACCTTCTCATCGTGC
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1)
15iBBGe(1,1)
2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
KC955130_BG8 GTTTCTTATTTGTTCCTTTTGCAGAGAAACACGTTCTAAAACTGGGTGAGTCCTCACTCCCAAATTATAAAGCAAAGGGTTCTGCCTGTGTGAGCTGTGG
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2) .....TTA.CCTC.....A
15iBBGb-2(2,1) .....TTA.CCTC.....A
15iBBGb-3(1,1) .....TTA.CCTC.....A
15iBBGb-4(1,1) .....TTA.CCTC.....A
15iBBGc(2,1) ...CC.TA.CCTC.....A
15iBBGd(1,1)
15iBBGe(1,1) .....TTA.CCTC.....A

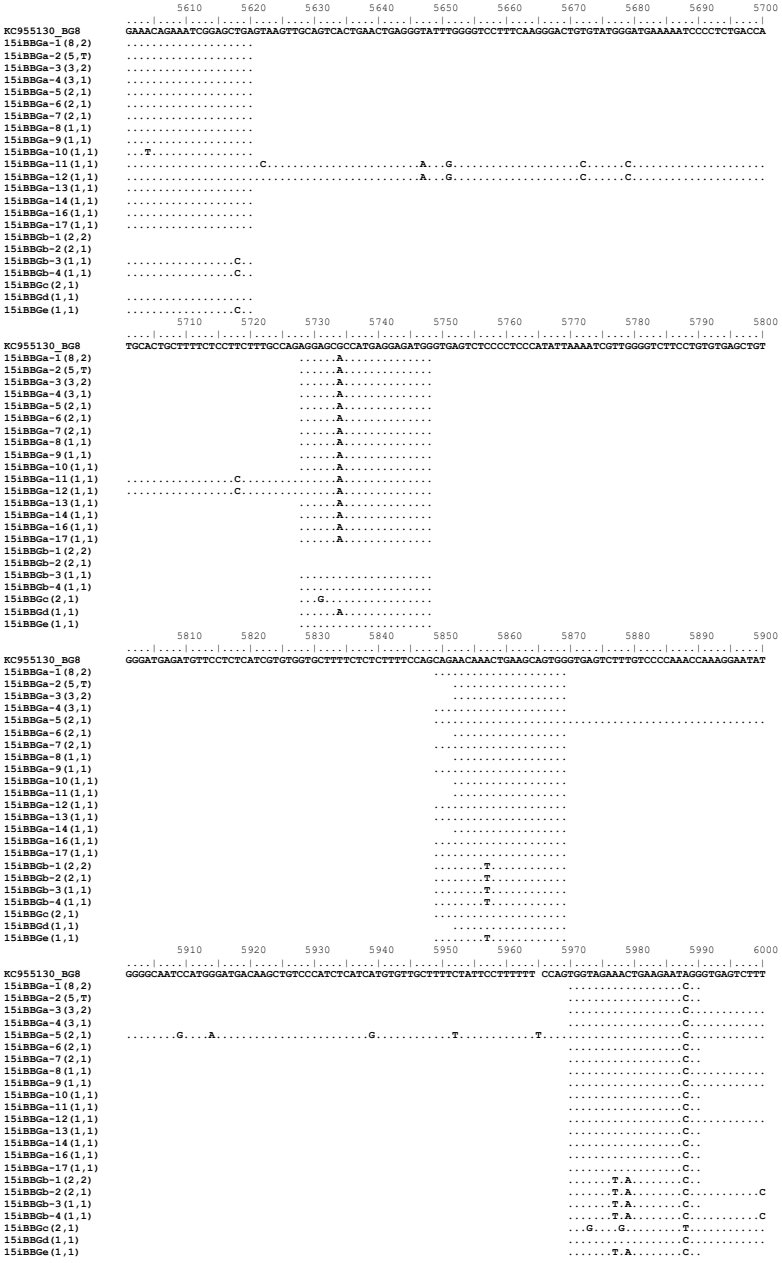

2810 2820 2830 2840 2850 2860 2870 2880 2890 2900
KC955130_BG8 GATCAGACGTTCCACTCATCATGCATTGCTTTTCTCTTTCTTTTTCAGAGGAAAAGACAGACGAAGTGGGTGAGTCTACATTCACTAAAGCAAAGAAATA
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1) ...G...T.....T......G.....
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2) .CA....TG.T.....
15iBBGb-2(2,1) .CA....TG.T.....
15iBBGb-3(1,1) .CA....TG.T.....
15iBBGb-4(1,1) .CA....TG.T.....
15iBBGc(2,1) .CAG..TTG.T.....
15iBBGd(1,1)
15iBBGe(1,1) .CA....TG.T.....
2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
KC955130_BG8 TGGGGTCTCCCATGGGATGACAAGCTGTCCCAAAAAATCATGTGGTGCTTTTTCTTGTCTTTTTATTTATTTATTTATTTATTTATTGCAGAGAATTGG
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2) ..G....C
15iBBGb-2(2,1) ..G....C
15iBBGb-3(1,1) ..G....C
15iBBGb-4(1,1) ..G....C
15iBBGc(2,1) ..G....C
15iBBGd(1,1)
15iBBGe(1,1) ..G....C
3010 3020 3030 3040 3050 3060 3070 3080 3090 3100
KC955130_BG8 AATTCAGTGCTGAGTAAGTTGCAGTCACTGAACTGAGGGAATGTGGGGTCTTCCTAAGGGACTGCGTAGGGGAGAAGTTCCCATGCACTGCTTTTCTCTT
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2) .....A.....
15iBBGb-2(2,1) .....A.....
15iBBGb-3(1,1) .....A.....
15iBBGb-4(1,1) .....A.....
15iBBGc(2,1) ....T..GT....
15iBBGd(1,1)
15iBBGe(1,1) .....A.....
3110 3120 3130 3140 3150 3160 3170 3180 3190 3200
KC955130_BG8 TCTTTTCCAGAGAAAGACAGTGAAGAGATGGGTGAGTCTCTCCTCCCAAATTAAAAACGTTGGGGTTCCCATGTGGGAGCTGTGGGATGAGATGTTCCTC
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2) ....G...T.....
15iBBGb-2(2,1) ....G...T.....
15iBBGb-3(1,1) ....G...T.....
15iBBGb-4(1,1) ....G...T.....
15iBBGc(2,1) ...G....T.....A....C
15iBBGd(1,1)
15iBBGe(1,1) ....G...T.....

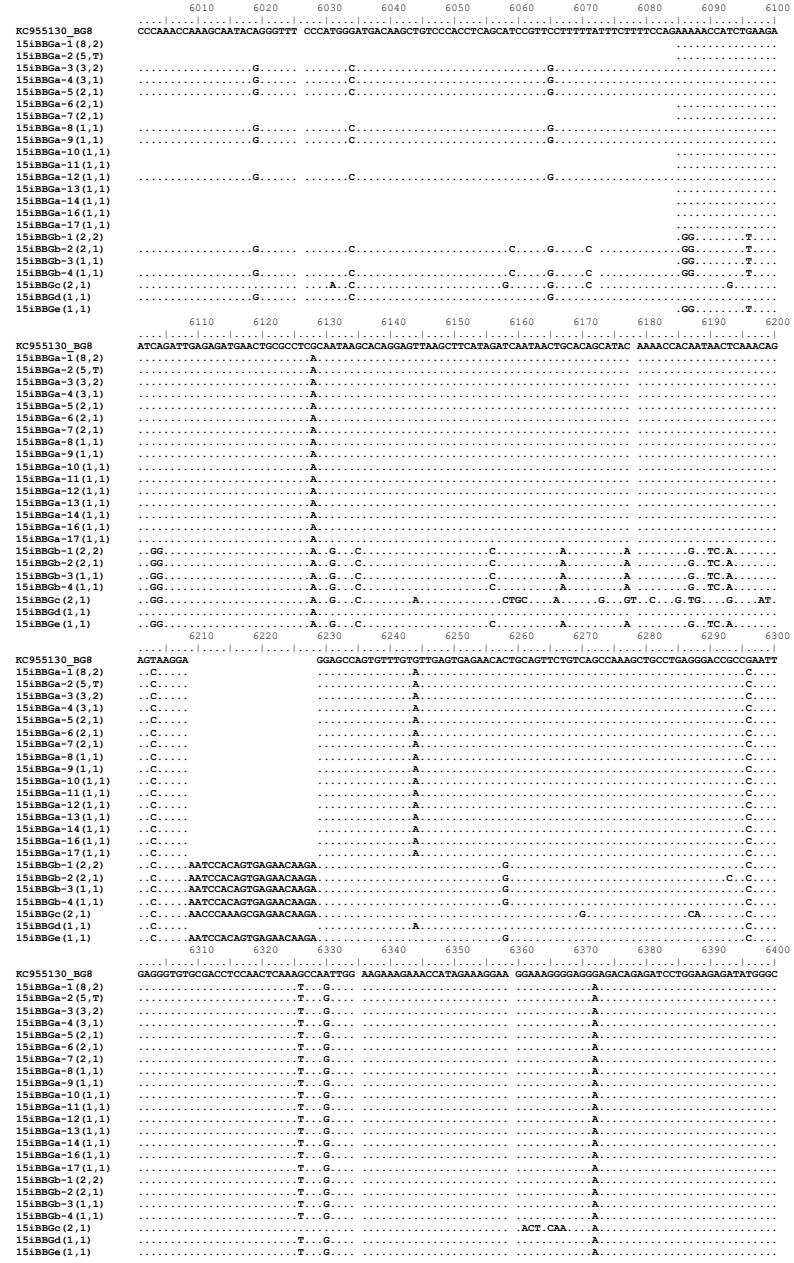

3210 3220 3230 3240 3250 3260 3270 3280 3290 3300
KC955130_BG8 TCATCAACCATCTTTTTTACTTTTCCTTTGCAGGTTATGGCTTTGCAGAACTGAGTAAGTCTCCCTCCCAACACGGAAGGGATTTGTGGTCTTCCCATGG
15iBBGa-1(8,2) ...C.....
15iBBGa-2(5,T) ...C.....
15iBBGa-3(3,2) ...C.....
15iBBGa-4(3,1) ...C.....
15iBBGa-5(2,1) ...C.....
15iBBGa-6(2,1) ...C.....
15iBBGa-7(2,1) ...C.....
15iBBGa-8(1,1) ...C.....
15iBBGa-9(1,1) ...C.....
15iBBGa-10(1,1) ...C.....
15iBBGa-11(1,1) ...C.....
15iBBGa-12(1,1) ...C.....
15iBBGa-13(1,1) ...C.....
15iBBGa-14(1,1) ...C.....
15iBBGa-16(1,1) ...C.....
15iBBGa-17(1,1) ...C.....
15iBBGb-1(2,2) ...C..T.G.....C....
15iBBGb-2(2,1) ...C..T.G.....C....
15iBBGb-3(1,1) ...C..T.G.....C....
15iBBGb-4(1,1) ...C..T.G.....C....
15iBBGc(2,1) ...C...G.....T....
15iBBGd(1,1) ...C.....
15iBBGe(1,1) ...C..T.G.....C....
3310 3320 3330 3340 3350 3360 3370 3380 3390 3400
KC955130_BG8 GATCAGCCATGGGATGATCATCTGACCCCTCTCATCATGCATTTCGTATTTGTTTCTTTGCAGAGAAACTGGCTGCAGAACTGGGTGAGTGCTGCCTCC
15iBBGa-1(8,2) .....A...A.....
15iBBGa-2(5,T) .....A...A.....
15iBBGa-3(3,2) .....A...A.....
15iBBGa-4(3,1) .....A...A.....
15iBBGa-5(2,1) .....A...A.....
15iBBGa-6(2,1) .....A...A.....
15iBBGa-7(2,1) .....A...A.....
15iBBGa-8(1,1) .....A...A.....
15iBBGa-9(1,1) .....A...A.....
15iBBGa-10(1,1) .....A...A.....
15iBBGa-11(1,1) .....A...A.....
15iBBGa-12(1,1) .....A...A.....
15iBBGa-13(1,1) .....A...A.....
15iBBGa-14(1,1) .....A...A.....
15iBBGa-16(1,1) .....A...A.....
15iBBGa-17(1,1) .....A...A.....
15iBBGb-1(2,2) .....T...
15iBBGb-2(2,1) .....T...
15iBBGb-3(1,1) .....T...
15iBBGb-4(1,1) .....T...
15iBBGc(2,1) ..G..G.....A.....
15iBBGd(1,1) .....A...A.....
15iBBGe(1,1) .....T...
3410 3420 3430 3440 3450 3460 3470 3480 3490 3500
KC955130_BG8 CAAATTAAATAAAAAATGGGTCTGCCTGGGAGAGTGGTGGGATGGCATGTTCCTCTCACTGCGTGTTGCTTTTCCTTTCTTTTCCAGAGAAACACTCTGA
15iBBGa-1(8,2) .....
15iBBGa-2(5,T) .....
15iBBGa-3(3,2) .....
15iBBGa-4(3,1) .....
15iBBGa-5(2,1) .....
15iBBGa-6(2,1) .....
15iBBGa-7(2,1) .....
15iBBGa-8(1,1) .....
15iBBGa-9(1,1) .....
15iBBGa-10(1,1) .....
15iBBGa-11(1,1) .....
15iBBGa-12(1,1) .....
15iBBGa-13(1,1) .....
15iBBGa-14(1,1) .....
15iBBGa-16(1,1) .....
15iBBGa-17(1,1) .....
15iBBGb-1(2,2) C.G.....A...
15iBBGb-2(2,1) C.G.....A...
15iBBGb-3(1,1) C.G.....A...
15iBBGb-4(1,1) C.G.....A...
15iBBGc(2,1) .....
15iBBGd(1,1) .....
15iBBGe(1,1) C.G.....A...
3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
KC955130_BG8 AGAGATGGGTGAGTCTCCCCTCCCCAATTATAAATGCTGGGGACTTCTTGTGGGAGCTGTGGGATGAGCTCTTCCTCTCATCATGCGCTGTTTCTGCTTT
15iBBGa-1(8,2) .....
15iBBGa-2(5,T) .....
15iBBGa-3(3,2) .....
15iBBGa-4(3,1) .....
15iBBGa-5(2,1) .....
15iBBGa-6(2,1) .....
15iBBGa-7(2,1) .....
15iBBGa-8(1,1) .....
15iBBGa-9(1,1) .....
15iBBGa-10(1,1) .....
15iBBGa-11(1,1) .....
15iBBGa-12(1,1) .....
15iBBGa-13(1,1) .....
15iBBGa-14(1,1) .....
15iBBGa-16(1,1) .....
15iBBGa-17(1,1) .....
15iBBGb-1(2,2) .....
15iBBGb-2(2,1) .....
15iBBGb-3(1,1) .....
15iBBGb-4(1,1) .....
15iBBGc(2,1) .....
15iBBGd(1,1) .....
15iBBGe(1,1) .....

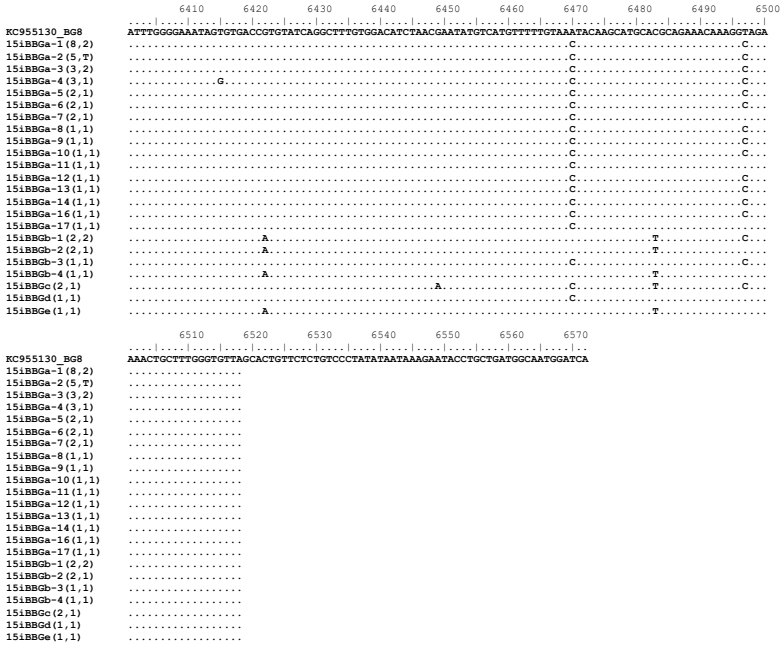

3610 3620 3630 3640 3650 3660 3670 3680 3690 3700
KC955130_BG8 TCCTTTGCAGGGACAAGGGATTTAAAGTTGGGTGAGTCTCTCTTCCCAAACCATACAGATTTGGGGTCTTCCCACGGCATCAGCCATGGGATGATAATCG
15iBBGa-1(8,2) ....G.....
15iBBGa-2(5,T) ....G.....
15iBBGa-3(3,2) ....G.....
15iBBGa-4(3,1) ....G.....
15iBBGa-5(2,1) ....G.....
15iBBGa-6(2,1) ....G.....
15iBBGa-7(2,1) ....G.....
15iBBGa-8(1,1) ....G.....
15iBBGa-9(1,1) ....G.....
15iBBGa-10(1,1) ....G.....
15iBBGa-11(1,1) ....G.....
15iBBGa-12(1,1) ....G.....
15iBBGa-13(1,1) ....G.....
15iBBGa-14(1,1) ....G.....
15iBBGa-16(1,1) ....G.....
15iBBGa-17(1,1) ....G.....
15iBBGb-1(2,2) ..CT.TT.AG.G...C....A
15iBBGb-2(2,1) ..CT.TT.AG.G...C....A
15iBBGb-3(1,1) ..CT.TT.AG.G...C....A
15iBBGb-4(1,1) ..CT.TT.AG.G...C....A
15iBBGc(2,1) .....T..GC...T....
15iBBGd(1,1) ....G.....
15iBBGe(1,1) ..CT.TT.AG.G...C....A
3710 3720 3730 3740 3750 3760 3770 3780 3790 3800
KC955130_BG8 GACCCTTCTCATCATGCATTTCTTATTGGTTCCTTTTGCAGAGCGACTAGCTGCCAAACTGGGTGAGTCCCCCCTCCCAAATTAAATAAAAAATGGGGTC
15iBBGa-1(8,2) .....
15iBBGa-2(5,T) .....
15iBBGa-3(3,2) .....
15iBBGa-4(3,1) .....
15iBBGa-5(2,1) .....
15iBBGa-6(2,1) .....
15iBBGa-7(2,1) .....
15iBBGa-8(1,1) .....
15iBBGa-9(1,1) .....
15iBBGa-10(1,1) .....
15iBBGa-11(1,1) .....
15iBBGa-12(1,1) .....
15iBBGa-13(1,1) .....
15iBBGa-14(1,1) .....
15iBBGa-16(1,1) .....
15iBBGa-17(1,1) .....
15iBBGb-1(2,2) ....A.....
15iBBGb-2(2,1) ....A.....
15iBBGb-3(1,1) ....A.....
15iBBGb-4(1,1) ....A.....
15iBBGc(2,1) ...T.....TG....A
15iBBGd(1,1) .....
15iBBGe(1,1) ....A.....
3810 3820 3830 3840 3850 3860 3870 3880 3890 3900
KC955130_BG8 TGCCTGTGTGAGCTGTGAGATGAGATGTTCCTCTCATCATGCGCTGCTTTTCTCTTCCTTTTCCAGAACATCAAACTAAAGAATTGGGTGAGTCTTCTTT
15iBBGa-1(8,2) ....A.G.....
15iBBGa-2(5,T) ....A.G.....
15iBBGa-3(3,2) ....A.G.....
15iBBGa-4(3,1) ....A.G.....
15iBBGa-5(2,1) ....A.G.....
15iBBGa-6(2,1) ....A.G.....
15iBBGa-7(2,1) ....A.G.....
15iBBGa-8(1,1) ....A.G.....
15iBBGa-9(1,1) ....A.G.....
15iBBGa-10(1,1) ....A.G.....
15iBBGa-11(1,1) ....A.G.....
15iBBGa-12(1,1) ....A.G.....
15iBBGa-13(1,1) ....A.G.....
15iBBGa-14(1,1) ....A.G.....
15iBBGa-16(1,1) ....A.G.....
15iBBGa-17(1,1) ....A.G.....
15iBBGb-1(2,2) ....A.G.....
15iBBGb-2(2,1) ....A.G.....
15iBBGb-3(1,1) ....A.G.....
15iBBGb-4(1,1) ....A.G.....
15iBBGc(2,1) G.CC...GA...CAAAGTCG. .TA.A.T.G..GCCA..C...
15iBBGd(1,1) ....A.G.....
15iBBGe(1,1) ....A.G....T.....
3910 3920 3930 3940 3950 3960 3970 3980 3990 4000
KC955130_BG8 CCCCAACCCCAAGAAATATGCGTTTCCCATGGGATGACAAGCTGTGCCACCTCATCATGCCCTGTTTTTTCTGTCCTTTTTGCAGAGAAACAGCATTCAC
15iBBGa-1(8,2) .C.....
15iBBGa-2(5,T) .C.....
15iBBGa-3(3,2) .C.....
15iBBGa-4(3,1) .C.....
15iBBGa-5(2,1) .C.....
15iBBGa-6(2,1) .C.....
15iBBGa-7(2,1) .C.....
15iBBGa-8(1,1) .C.....
15iBBGa-9(1,1) .C.....
15iBBGa-10(1,1) .C.....
15iBBGa-11(1,1) .C.....
15iBBGa-12(1,1) .C.....
15iBBGa-13(1,1) .C.....
15iBBGa-14(1,1) .C.....
15iBBGa-16(1,1) .C.....
15iBBGa-17(1,1) .C.....
15iBBGb-1(2,2) .C.....A.....
15iBBGb-2(2,1) .C.....A.....
15iBBGb-3(1,1) .C.....A.....
15iBBGb-4(1,1) .C.....A.....
15iBBGc(2,1) ..CT.....
15iBBGd(1,1) .C.....
15iBBGe(1,1) .C.....A.....


4010 4020 4030 4040 4050 4060 4070 4080 4090 4100
KC955130_BG8 AGTTCCGTAAGTTGCAGTCACTGAACTGAAGGAATGTGGGGTCTTCCCAAAGTCCTGCATGTGGGATGAAAAATCCCCTCTGACCATGCACTGCTTTTCT
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1)
15iBBGe(1,1)
4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
KC955130_BG8 CCTTCTATTCCAGAGAGACACTTTCAGAATATGTGTGAGTCTCCCCACCCCTGATAAATAAAAACGTTGGGGTCTTGCTGTGTGAGCTGTGGGATGAGAT
15iBBGa-1(8,2) .C.....
15iBBGa-2(5,T) .C.....
15iBBGa-3(3,2) .C.....
15iBBGa-4(3,1) .C.....
15iBBGa-5(2,1) .C.....
15iBBGa-6(2,1) .C.....
15iBBGa-7(2,1) .C.....
15iBBGa-8(1,1) .C.....
15iBBGa-9(1,1) .C.....
15iBBGa-10(1,1) .C.....
15iBBGa-11(1,1) .C.....
15iBBGa-12(1,1) .C.....
15iBBGa-13(1,1) .C.....
15iBBGa-14(1,1) .C.....
15iBBGa-16(1,1) .C.....
15iBBGa-17(1,1) .C.....
15iBBGb-1(2,2) .....G
15iBBGb-2(2,1) .....G
15iBBGb-3(1,1) .....G
15iBBGb-4(1,1) .....G
15iBBGc(2,1) .....G
15iBBGd(1,1) .C.....
15iBBGe(1,1) .....G
4210 4220 4230 4240 4250 4260 4270 4280 4290 4300
KC955130_BG8 GTTCCTCTCATCACACATTGTTTTTCTGATTTTCTTTTGCAGATTTAAGTGCTGGAAAACAGAGTAAGTCTCCCTCCCTGCACAGAAGGAACTTACGGTT
15iBBGa-1(8,2) .....
15iBBGa-2(5,T) .....
15iBBGa-3(3,2) .....
15iBBGa-4(3,1) .....
15iBBGa-5(2,1) .....
15iBBGa-6(2,1) .....
15iBBGa-7(2,1) .....
15iBBGa-8(1,1) .....
15iBBGa-9(1,1) .....
15iBBGa-10(1,1) .....
15iBBGa-11(1,1) .....
15iBBGa-12(1,1) .....
15iBBGa-13(1,1) .....
15iBBGa-14(1,1) .....
15iBBGa-16(1,1) .....
15iBBGa-17(1,1) .....
15iBBGb-1(2,2) ....C.....
15iBBGb-2(2,1) ....C.....
15iBBGb-3(1,1) ....C.....
15iBBGb-4(1,1) ....C.....
15iBBGc(2,1) .....
15iBBGd(1,1) .....
15iBBGe(1,1) ....C.....
4310 4320 4330 4340 4350 4360 4370 4380 4390 4400
KC955130_BG8 TTCCCATGGGATCAGCCATGGGATCATCATCCGACTCTTCTCATCATGAATTTCGTCTTTCTTTCTTTTGCAGAGAAAATGGTTACAAAACTGGGTGAGT
15iBBGa-1(8,2) .....
15iBBGa-2(5,T) .....
15iBBGa-3(3,2) .....
15iBBGa-4(3,1) .....
15iBBGa-5(2,1) .....
15iBBGa-6(2,1) .....
15iBBGa-7(2,1) .....
15iBBGa-8(1,1) .....
15iBBGa-9(1,1) .....
15iBBGa-10(1,1) .....
15iBBGa-11(1,1) .....
15iBBGa-12(1,1) .....
15iBBGa-13(1,1) .....
15iBBGa-14(1,1) .....
15iBBGa-16(1,1) .....
15iBBGa-17(1,1) .....
15iBBGb-1(2,2) ..G..C.....
15iBBGb-2(2,1) ..G..C.....
15iBBGb-3(1,1) ..G..C.....
15iBBGb-4(1,1) ..G..C.....
15iBBGc(2,1) ..G..C.....T...
15iBBGd(1,1) .....
15iBBGe(1,1) ..G..C.....
4410 4420 4430 4440 4450 4460 4470 4480 4490 4500
KC955130_BG8 CCAACCTCCCAAACTAAATTAAAAACAGTCAGACTTTGTGAGCTGTGGGATGAGACGTTCCTCTCATCATGTGCTGCTTTCCTTTTACTTTTCCAGAGGA
15iBBGa-1(8,2) ....
15iBBGa-2(5,T) ....
15iBBGa-3(3,2) ....
15iBBGa-4(3,1) ....
15iBBGa-5(2,1) ....
15iBBGa-6(2,1) ....
15iBBGa-7(2,1) ....
15iBBGa-8(1,1) ....
15iBBGa-9(1,1) ....
15iBBGa-10(1,1) ....
15iBBGa-11(1,1) ....
15iBBGa-12(1,1) ....
15iBBGa-13(1,1) ....
15iBBGa-14(1,1) ....
15iBBGa-16(1,1) ....
15iBBGa-17(1,1) ....
15iBBGb-1(2,2) ....
15iBBGb-2(2,1) ....
15iBBGb-3(1,1) ....
15iBBGb-4(1,1) ....
15iBBGc(2,1) ....
15iBBGd(1,1) ....
15iBBGe(1,1) ....
4510 4520 4530 4540 4550 4560 4570 4580 4590 4600
KC955130_BG8 ACACTGTGAATGGATGGGTGAGTCTCCCCTCCCAAATTAAAAATGTTGGGGTCTTCCTGTGAGAGCTGTGGGATGAGCTGTTCCTCTCATCGTGCACTGT
15iBBGa-1(8,2) .....TGAGAAGGAATG.AAATTTGG.GGC.GCAGCT.TAAAAGTG.
15iBBGa-2(5,T) .....TGAGAAGGAATG.AAATTTGG.GGC.GCAGCT.TAAAAGTG.
15iBBGa-3(3,2) .....TGAGAAGGAATG.AAATTTGG.GGC.GCAGCT.TAAAAGTG.
15iBBGa-4(3,1) .....TGAGAAGGAATG.AAATTTGG.GGC.GCAGCT.TAAAAGTG.
15iBBGa-5(2,1) .....TGAGAAGGAATG.AAATTTGG.GGC.GCAGCT.TAAAAGTG.
15iBBGa-6(2,1) .....TGAGAAGGAATG.AAATTTGG.GGC.GCAGCT.TAAAAGTG.
15iBBGa-7(2,1) .....TGAGAAGGAATG.AAATTTGG.GGC.GCAGCT.TAAAAGTG.
15iBBGa-8(1,1) .....TGAGAAGGAATG.AAATTTGG.GGC.GCAGCT.TAAAAGTG.
15iBBGa-9(1,1) .....TGAGAAGGAATG.AAATTTGG.GGC.GCAGCT.TAAAAGTG.
15iBBGa-10(1,1) .....TGAGAAGGAATG.AAATTTGG.GGC.GCAGCT.TAAAAGTG.
15iBBGa-11(1,1) .....TGAGAAGGAATG.AAATTTGG.GGC.GCAGCT.TAAAAGTG.
15iBBGa-12(1,1) .....TGAGAAGGAATG.AAATTTGG.GGC.GCAGCT.TAAAAGTG.
15iBBGa-13(1,1) .....TGAGAAGGAATG.AAATTTGG.GGC.GCAGCT.TAAAAGTG.
15iBBGa-14(1,1) .....TGAGAAGGAATG.AAATTTGG.GGC.GCAGCT.TAAAAGTG.
15iBBGa-16(1,1) .....TGAGAAGGAATG.AAATTTGG.GGC.GCAGCT.TAAAAGTG.
15iBBGa-17(1,1) .....TGAGAAGGAATG.AAATTTGG.GGC.GCAGCT.TAAAAGTG.
15iBBGb-1(2,2) .....CA.....
15iBBGb-2(2,1) .....CA.....
15iBBGb-3(1,1) .....CA.....
15iBBGb-4(1,1) .....CA.....
15iBBGc(2,1) C.....C.....
15iBBGd(1,1) .....TGAGAAGGAATG.AAATTTGG.GGC.GCAGCT.TAAAAGTG.
15iBBGe(1,1) .....CA.....
4610 4620 4630 4640 4650 4660 4670 4680 4690 4700
KC955130_BG8 TTCTGCTTTTCCTTTGCAGTGAGAAGGAATGTAAAGTTGGGTGAGTCTTCTTCCCCAACCAAAGAGATTCGGAGTCTTCCATGGGATCAGCCATGGGATG
15iBBGa-1(8,2) .....
15iBBGa-2(5,T) .....
15iBBGa-3(3,2) .....
15iBBGa-4(3,1) .....
15iBBGa-5(2,1) .....
15iBBGa-6(2,1) .....
15iBBGa-7(2,1) .....
15iBBGa-8(1,1) .....
15iBBGa-9(1,1) .....
15iBBGa-10(1,1) .....
15iBBGa-11(1,1) .....
15iBBGa-12(1,1) .....
15iBBGa-13(1,1) .....
15iBBGa-14(1,1) .....
15iBBGa-16(1,1) .....
15iBBGa-17(1,1) .....
15iBBGb-1(2,2) A.....
15iBBGb-2(2,1) A.....
15iBBGb-3(1,1) A.....
15iBBGb-4(1,1) A.....
15iBBGc(2,1) .....
15iBBGd(1,1) .....
15iBBGe(1,1) A.....
4710 4720 4730 4740 4750 4760 4770 4780 4790 4800
KC955130_BG8 ATAACATGAACCTCATCACGTGTTTCTTATTTGTTCCTTTTGCAGAGGCAGCAGCTGTAAAAGTGGGTGAGTCTCCCCTCCCAAATTAAAAATGTTGGCG
15iBBGa-1(8,2) .....
15iBBGa-2(5,T) .....
15iBBGa-3(3,2) .....
15iBBGa-4(3,1) .....
15iBBGa-5(2,1) .....
15iBBGa-6(2,1) .....
15iBBGa-7(2,1) .....
15iBBGa-8(1,1) .....
15iBBGa-9(1,1) .....
15iBBGa-10(1,1) .....
15iBBGa-11(1,1) .....
15iBBGa-12(1,1) .....
15iBBGa-13(1,1) .....
15iBBGa-14(1,1) .....
15iBBGa-16(1,1) .....
15iBBGa-17(1,1) .....
15iBBGb-1(2,2) .....C.....
15iBBGb-2(2,1) .....C.....
15iBBGb-3(1,1) .....C.....
15iBBGb-4(1,1) .....C.....
15iBBGc(2,1) ..AT.C.....
15iBBGd(1,1) .....
15iBBGe(1,1) .....C.....

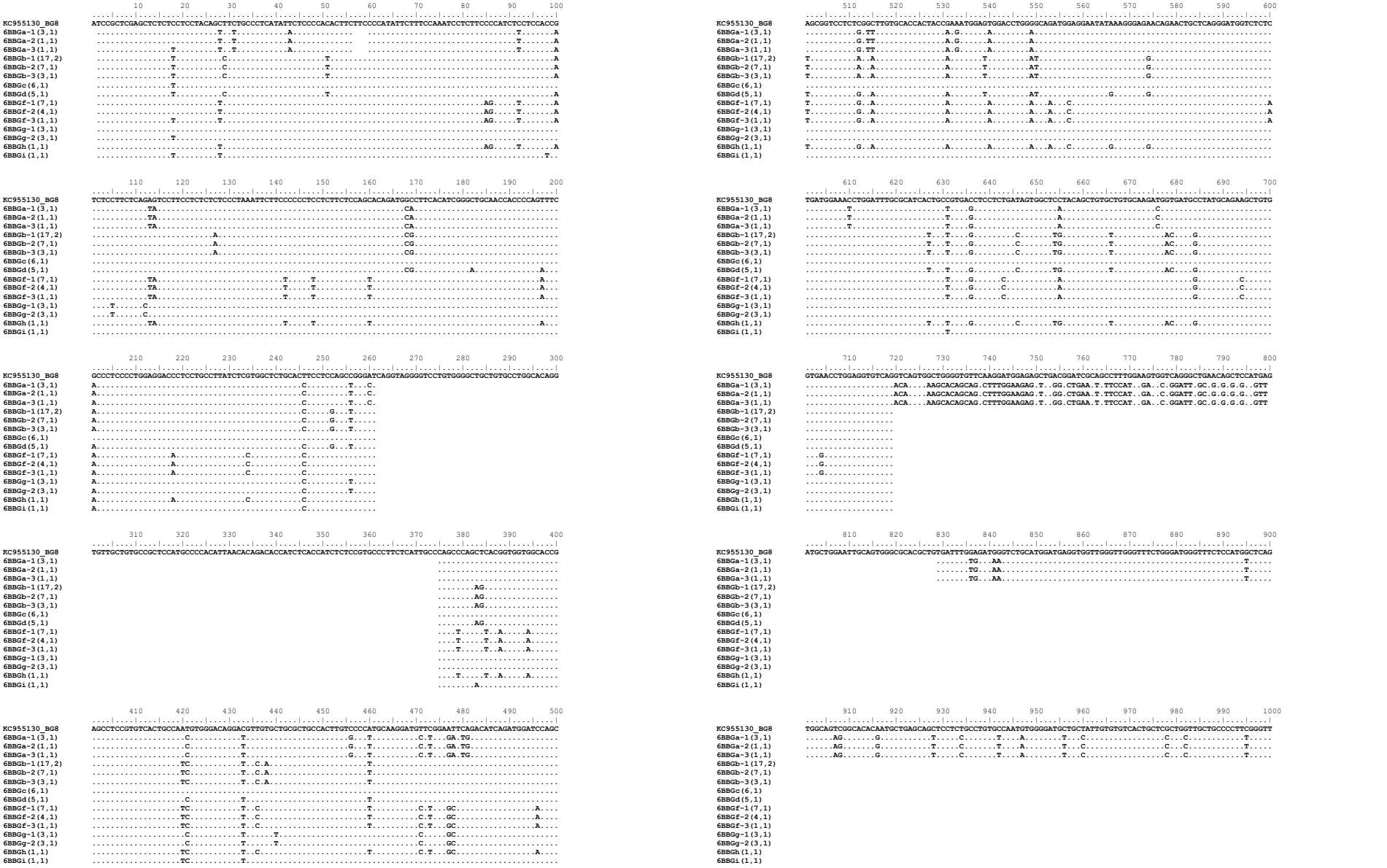

4810 4820 4830 4840 4850 4860 4870 4880 4890 4900
KC955130_BG8 TCATCCTGTGAGAGCTGTGGGATGAGCTGTTCCTCTCATCGTGCACTGTTTCTGCTTTTCCTTTGCAGTGAGAAGGAATGTAAAGTTGGGTGAGTCTTCT
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1)
15iBBGe(1,1)
4910 4920 4930 4940 4950 4960 4970 4980 4990 5000
KC955130_BG8 TCCCCAACCAAAGAGATGTGGGGTCTTCCATGGGATCAGCCATGGGATGATAAGCTGAACCTTATCACGTGTTTCTTATTTGTTCCTTTTGCAGAGGCAG
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1)
15iBBGe(1,1)
5010 5020 5030 5040 5050 5060 5070 5080 5090 5100
KC955130_BG8 CAGCTGTAAAAGTGGGTGAGTCCTCCCTCCCAAATCAAATACAAAAGGGGATCTGCCTGTGTGAGCTGTGGGATGAGATGTTCCTCTCATCACGCATTGT
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1)
15iBBGe(1,1)
5110 5120 5130 5140 5150 5160 5170 5180 5190 5200
KC955130_BG8 TTTTCTCATTCATTTCCAGGACACAAAGCTAAAGAATCAGGTGAGTCTTCTTCCCTGTCCCAAAGGACTATGGGTTTCCCATGGGATGACAAGCTGTGCC
15iBBGa-1(8,2) .....
15iBBGa-2(5,T) .....
15iBBGa-3(3,2) .....
15iBBGa-4(3,1) .....
15iBBGa-5(2,1) .....
15iBBGa-6(2,1) .....
15iBBGa-7(2,1) .....A.....
15iBBGa-8(1,1) .....
15iBBGa-9(1,1) .....
15iBBGa-10(1,1) .....
15iBBGa-11(1,1) .....
15iBBGa-12(1,1) .....
15iBBGa-13(1,1) .....C..A.....G.....T.....
15iBBGa-14(1,1) .....
15iBBGa-16(1,1) .....A.....
15iBBGa-17(1,1) .....
15iBBGb-1(2,2) .G..AC.....
15iBBGb-2(2,1) .G..AC.....C...
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1) .....
15iBBGe(1,1)
5210 5220 5230 5240 5250 5260 5270 5280 5290 5300
KC955130_BG8 ACCTCCTCATGAGGTGCTTCTTCTTTCTTTGTGCAGAGAAACAGAAATCGGAGCTGAGTAAGTTGCAGTCACTGAACTGAGGGAATGTGGGGTCTTCCCA
15iBBGa-1(8,2) .....
15iBBGa-2(5,T) .....
15iBBGa-3(3,2) .....
15iBBGa-4(3,1) .....
15iBBGa-5(2,1) .....
15iBBGa-6(2,1) .....A.....
15iBBGa-7(2,1) .....
15iBBGa-8(1,1) .....
15iBBGa-9(1,1) .....
15iBBGa-10(1,1) .....
15iBBGa-11(1,1) .....A.....
15iBBGa-12(1,1) .....A.....
15iBBGa-13(1,1) .....T.....A.....
15iBBGa-14(1,1) .....
15iBBGa-16(1,1) .....
15iBBGa-17(1,1) .....
15iBBGb-1(2,2) .....C..
15iBBGb-2(2,1) .....C..
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1) .....
15iBBGe(1,1)
5310 5320 5330 5340 5350 5360 5370 5380 5390 5400
KC955130_BG8 AAGTCCTGCGTATGGGATGAAAAATCCCCTCTGACCATGCACTGCTTTTCTCCTCCTTTGCCAGAGGAGCGCCATGAGGAGATGGGTGAGTCTCCCCTCC
15iBBGa-1(8,2) .....
15iBBGa-2(5,T) .....
15iBBGa-3(3,2) .....
15iBBGa-4(3,1) .....
15iBBGa-5(2,1) .....
15iBBGa-6(2,1) .....
15iBBGa-7(2,1) .....
15iBBGa-8(1,1) .....
15iBBGa-9(1,1) .....
15iBBGa-10(1,1) .....
15iBBGa-11(1,1) .....
15iBBGa-12(1,1) .....
15iBBGa-13(1,1) .....
15iBBGa-14(1,1) .....
15iBBGa-16(1,1) .....
15iBBGa-17(1,1) .....
15iBBGb-1(2,2) .....
15iBBGb-2(2,1) .....
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1) .....
15iBBGe(1,1)
5410 5420 5430 5440 5450 5460 5470 5480 5490 5500
KC955130_BG8 CATATTAAAATCGTTGGGGTCTTCCTGTGTGAGCTGTGGGATGAGATGTTCCTCTCATCACACATTGTTTTTCTTTTCCAGGGCAACAAGCTAAAGAATC
15iBBGa-1(8,2) .....
15iBBGa-2(5,T) .....
15iBBGa-3(3,2) .....
15iBBGa-4(3,1) .....
15iBBGa-5(2,1) .....
15iBBGa-6(2,1) .....
15iBBGa-7(2,1) .....
15iBBGa-8(1,1) .....
15iBBGa-9(1,1) .....
15iBBGa-10(1,1) .....
15iBBGa-11(1,1) .....
15iBBGa-12(1,1) .....
15iBBGa-13(1,1) .....
15iBBGa-14(1,1)
15iBBGa-16(1,1) .....
15iBBGa-17(1,1) .....
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1) .....
15iBBGb-4(1,1) .....
15iBBGc(2,1) .....
15iBBGd(1,1) .....
15iBBGe(1,1) .....
5510 5520 5530 5540 5550 5560 5570 5580 5590 5600
KC955130_BG8 AGGTGAGTCTTCTTCCCCGTCCCAAAGGACTATGGGTTTCCCATGGGATGACAAGCTGTGCCACCTCCTCATGAGGTGCTTCTTCTTTCTTTTTTGCAGA
15iBBGa-1(8,2) .. .
15iBBGa-2(5,T) .. .
15iBBGa-3(3,2) .. .
15iBBGa-4(3,1) .. .
15iBBGa-5(2,1) .. .
15iBBGa-6(2,1) .. .
15iBBGa-7(2,1) .. .
15iBBGa-8(1,1) .. .
15iBBGa-9(1,1) .. .
15iBBGa-10(1,1) .
15iBBGa-11(1,1) .. .
15iBBGa-12(1,1) .. .
15iBBGa-13(1,1) .. .
15iBBGa-14(1,1) .
15iBBGa-16(1,1) .. .
15iBBGa-17(1,1) .. .
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1) .. .
15iBBGb-4(1,1) .. .
15iBBGc(2,1) G.
15iBBGd(1,1) .. .
15iBBGe(1,1) .. .

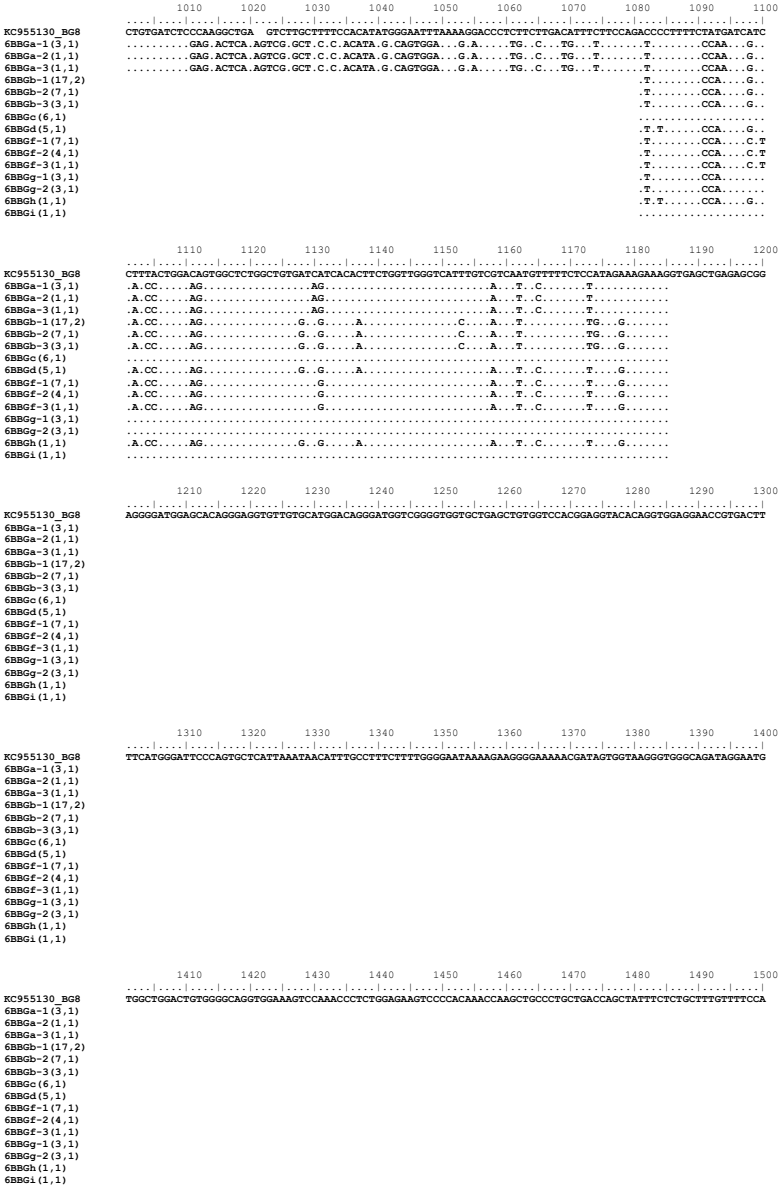

5610 5620 5630 5640 5650 5660 5670 5680 5690 5700
KC955130_BG8 GAAACAGAAATCGGAGCTGAGTAAGTTGCAGTCACTGAACTGAGGGTATTTGGGGTCCTTTCAAGGGACTGTGTATGGGATGAAAAATCCCCTCTGACCA
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1)
15iBBGe(1,1)
5710 5720 5730 5740 5750 5760 5770 5780 5790 5800
KC955130_BG8 TGCACTGCTTTTCTCCTTCTTTGCCAGAGGAGCGCCATGAGGAGATGGGTGAGTCTCCCCCTCCCATATTAAAATCGTTGGGGTCTTCCTGTGTGAGCTGT
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1)
15iBBGe(1,1)
5810 5820 5830 5840 5850 5860 5870 5880 5890 5900
KC955130_BG8 GGGATGAGATGTTCCTCTCATCGTGTGGTGCTTTTCTCTCTTTTCCAGCAGAACAAACTGAAGCAGTGGGTGAGTCTTTGTCCCCAAACCAAAGGAATAT
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1)
15iBBGe(1,1)
5910 5920 5930 5940 5950 5960 5970 5980 5990 6000
KC955130_BG8 GGGGCAATCCATGGGATGACAAGCTGTCCCATCTCATCATGTGTTGCTTTTCTATTCCTTTTTT CCAGTGGTAGAAACTGAAGAATAGGGTGAGTCTTT
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1)
15iBBGe(1,1)

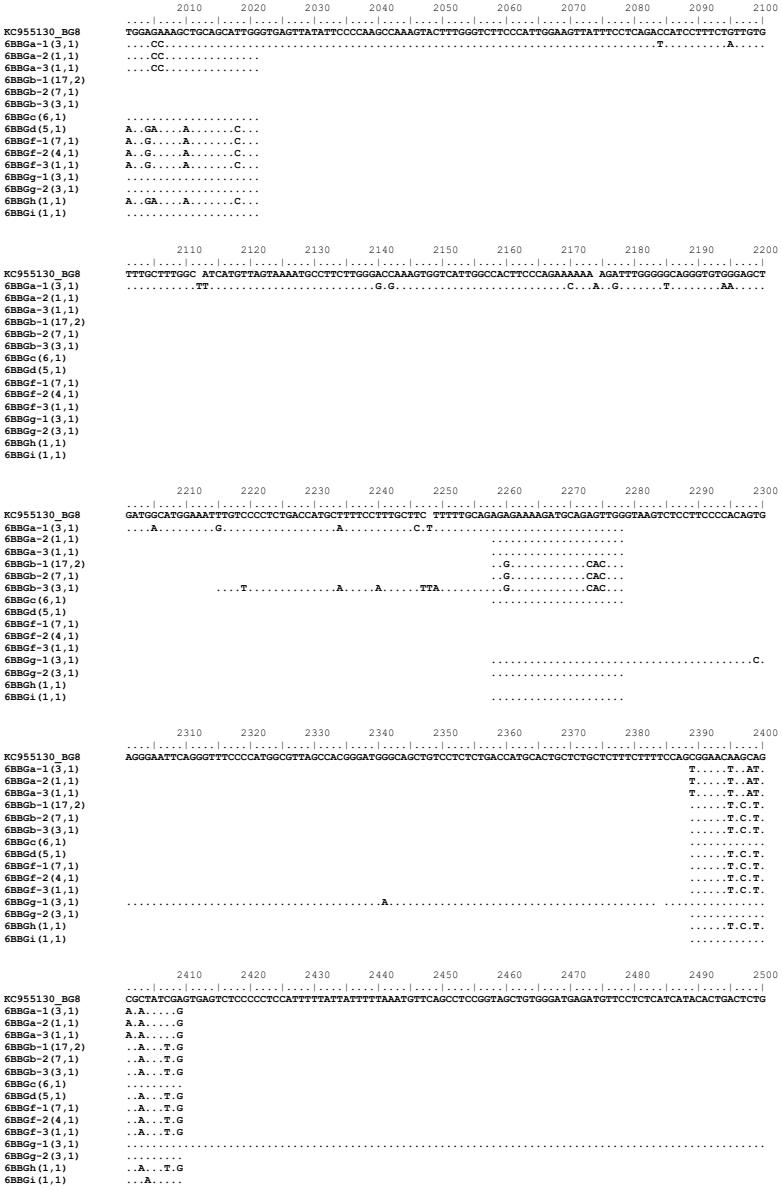

6010 6020 6030 6040 6050 6060 6070 6080 6090 6100
KC955130_BG8 CCCAAACCAAAGCAATACAGGGTTT CCCATGGGATGACAAGCTGTCCCACCTCAGCATCCGTTCCTTTTTATTTCTTTTCCAGAAAAACCATCTGAAGA
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1)
15iBBGe(1,1)
6110 6120 6130 6140 6150 6160 6170 6180 6190 6200
KC955130_BG8 ATCAGATTGAGAGATGAACTGCGCCTCGCAATAAGCACAGGAGTTAAGCTTCATAGATCAATAACTGCACAGCATAC AAAACCACAATAACTCAAACAG
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1)
15iBBGe(1,1)
6210 6220 6230 6240 6250 6260 6270 6280 6290 6300
KC955130_BG8 AGTAAGGA GGAGCCAGTGTTTGTGTTGAGTGAGAACACTGCAGTTCTGTCAGCCAAAGCTGCCTGAGGGACCGCCGAATT
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2) AATCCACAGTGAGAACAAGA
15iBBGb-2(2,1) AATCCACAGTGAGAACAAGA
15iBBGb-3(1,1) AATCCACAGTGAGAACAAGA
15iBBGb-4(1,1) AATCCACAGTGAGAACAAGA
15iBBGc(2,1) AACCCAAAGCGAGAACAAGA
15iBBGd(1,1)
15iBBGe(1,1) AATCCACAGTGAGAACAAGA
6310 6320 6330 6340 6350 6360 6370 6380 6390 6400
KC955130_BG8 GAGGGTGTGCGACCTCCAACTCAAAGCCAATTGG AAGAAAGAAACCATAGAAAGGAA GGAAAGGGGAGGGAGACAGAGATCCTGGAAGAGATATGGGC
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1)
15iBBGe(1,1)

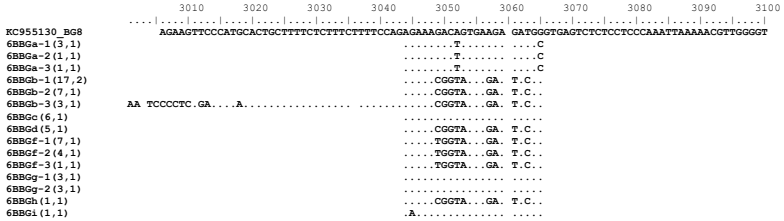

6410 6420 6430 6440 6450 6460 6470 6480 6490 6500
KC955130_BG8 ATTTGGGGAAATAGTGTGACCGTGTATCAGGCTTTGTGGACATCTAACGAATATGTCATGTTTTTGTAAATACAAGCATGCACGCAGAAACAAAGGTAGA
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1)
15iBBGe(1,1)
6510 6520 6530 6540 6550 6560 6570
KC955130_BG8 AAACTGCTTTGGGTGTTAGCACTGTTCTCTGTCCCTATATAATAAAGAATACCTGCTGATGGCAATGGATCA
15iBBGa-1(8,2)
15iBBGa-2(5,T)
15iBBGa-3(3,2)
15iBBGa-4(3,1)
15iBBGa-5(2,1)
15iBBGa-6(2,1)
15iBBGa-7(2,1)
15iBBGa-8(1,1)
15iBBGa-9(1,1)
15iBBGa-10(1,1)
15iBBGa-11(1,1)
15iBBGa-12(1,1)
15iBBGa-13(1,1)
15iBBGa-14(1,1)
15iBBGa-16(1,1)
15iBBGa-17(1,1)
15iBBGb-1(2,2)
15iBBGb-2(2,1)
15iBBGb-3(1,1)
15iBBGb-4(1,1)
15iBBGc(2,1)
15iBBGd(1,1)
15iBBGe(1,1)

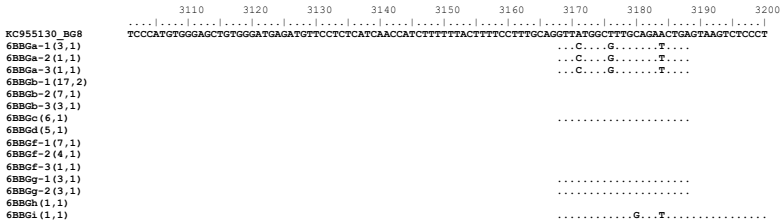
(J4)

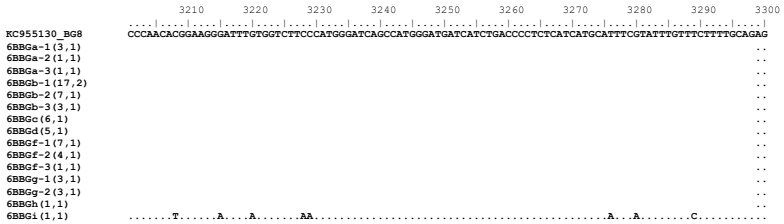

KC955130_BG8
6BBGa-1(3,1)
6BBGa-2(1,1)
6BBGa-3(1,1)
6BBGb-1(17,2)
6BBGb-2(7,1)
6BBGb-3(3,1)
6BBGc(6,1)
6BBGd(5,1)
6BBGf-1(7,1)
6BBGf-2(4,1)
6BBGf-3(1,1)
6BBGg-1(3,1)
6BBGg-2(3,1)
6BBGh(1,1)
6BBGi(1,1)

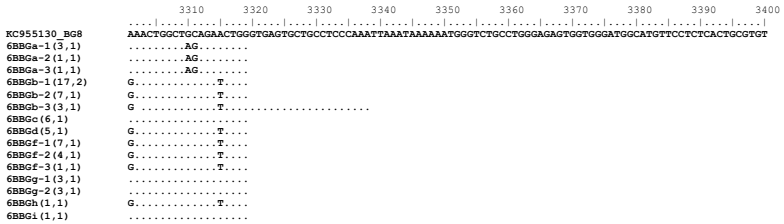

1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
KC955130_BG8 CTGTGATCTCCCAAGGCTGA GTCTTGCTTTTCCACATATGGGAATTTAAAAGGACCCTCTTCTTGACATTTCTTCCAGACCCCTTTTCTATGATCATC
6BBGa-1(3,1) ..........GAG.ACTCA.AGTCG.GCT.C.C.ACATA.G.CAGTGGA...G.A.....TG..C...TG...T.......T.........CCAA...G..
6BBGa-2(1,1) ..........GAG.ACTCA.AGTCG.GCT.C.C.ACATA.G.CAGTGGA...G.A.....TG..C...TG...T.......T.........CCAA...G..
6BBGa-3(1,1) ..........GAG.ACTCA.AGTCG.GCT.C.C.ACATA.G.CAGTGGA...G.A.....TG..C...TG...T.......T.........CCAA...G..
6BBGb-1(17,2) .T.........CCA....G..
6BBGb-2(7,1) .T.........CCA....G..
6BBGb-3(3,1) .T.........CCA....G..
6BBGc(6,1) ....
6BBGd(5,1) .T.T......CCA....G..
6BBGf-1(7,1) .T.........CCA....C.T
6BBGf-2(4,1) .T.........CCA....C.T
6BBGf-3(1,1) .T.........CCA....C.T
6BBGg-1(3,1) .T.........CCA.......
6BBGg-2(3,1) .T.........CCA.......
6BBGh(1,1) .T.T......CCA....G..
6BBGi(1,1) ....
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
KC955130_BG8 CTTTACTGGACAGTGGCTCTGGCTGTGATCATCACACTTCTGGTTGGGTCATTTGTCGTCAATGTTTTTCTCCATAGAAAGAAAGGTGAGCTGAGAGCGG
6BBGa-1(3,1) .A.CC.....AG.....................AG.....................A...T..C.....T...........
6BBGa-2(1,1) .A.CC.....AG.....................AG.....................A...T..C.....T...........
6BBGa-3(1,1) .A.CC.....AG.....................AG.....................A...T..C.....T...........
6BBGb-1(17,2) .A.CC.....AG.....G..G....A.....C....A...T.....TG...G.....
6BBGb-2(7,1) .A.CC.....AG.....G..G....A.....C....A...T.....TG...G.....
6BBGb-3(3,1) .A.CC.....AG.....G..G....A.....C....A...T.....TG...G.....
6BBGc(6,1) ....
6BBGd(5,1) .A.CC.....AG.....G..G....A.....A...T..C.....T....G.....
6BBGf-1(7,1) .A.CC.....AG.....G.....A...T..C.....T....G.....
6BBGf-2(4,1) .A.CC.....AG.....G.....A...T..C.....T....G.....
6BBGf-3(1,1) .A.CC.....AG.....G.....A...T..C.....T....G.....
6BBGg-1(3,1) ....
6BBGg-2(3,1) ....
6BBGh(1,1) .A.CC.....AG.....G..G....A.....A...T..C.....T....G.....
6BBGi(1,1) ....
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
KC955130_BG8 AGGGGATGGAGCACAGGGAGGTGTTGTGCATGGACAGGGATGGTCGGGGTGGTGCTGAGCTGTGGTCCACGGAGGTACACAGGTGGAGGAACCGTGACTT
6BBGa-1(3,1)
6BBGa-2(1,1)
6BBGa-3(1,1)
6BBGb-1(17,2)
6BBGb-2(7,1)
6BBGb-3(3,1)
6BBGc(6,1)
6BBGd(5,1)
6BBGf-1(7,1)
6BBGf-2(4,1)
6BBGf-3(1,1)
6BBGg-1(3,1)
6BBGg-2(3,1)
6BBGh(1,1)
6BBGi(1,1)
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
KC955130_BG8 TTCATGGGATTCCCAGTGCTCATTAAATAACATTTGCCTTTCTTTTGGGGAATAAAAGAAGGGGAAAAACGATAGTGGTAAGGGTGGGCAGATAGGAATG
6BBGa-1(3,1)
6BBGa-2(1,1)
6BBGa-3(1,1)
6BBGb-1(17,2)
6BBGb-2(7,1)
6BBGb-3(3,1)
6BBGc(6,1)
6BBGd(5,1)
6BBGf-1(7,1)
6BBGf-2(4,1)
6BBGf-3(1,1)
6BBGg-1(3,1)
6BBGg-2(3,1)
6BBGh(1,1)
6BBGi(1,1)
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
KC955130_BG8 TGGCTGGACTGTGGGGCAGGTGGAAAGTCCAAACCCTCTGGAGAAGTCCCCACAAACCAAGCTGCCCTGCTGACCAGCTATTTCTCTGCTTTGTTTTCCA
6BBGa-1(3,1)
6BBGa-2(1,1)
6BBGa-3(1,1)
6BBGb-1(17,2)
6BBGb-2(7,1)
6BBGb-3(3,1)
6BBGc(6,1)
6BBGd(5,1)
6BBGf-1(7,1)
6BBGf-2(4,1)
6BBGf-3(1,1)
6BBGg-1(3,1)
6BBGg-2(3,1)
6BBGh(1,1)
6BBGi(1,1)
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
KC955130_BG8 GTGGCACAGAGCAGAGAGCTGAGTGAGTCCTTCCATCCCCATCCACCACCAAAGTCCCTTTAATGGAACTGACAGCAGACTGCAGAGTGCTGGGTT ATG
6BBGa-1(3,1) C.....................................G.T..T.....................................T..A.....................T...
6BBGa-2(1,1) C.....................................G.T..T.....................................T..A.....................T...
6BBGa-3(1,1) C.....................
6BBGb-1(17,2) ...T.................
6BBGb-2(7,1) ...T.................
6BBGb-3(3,1) ...T.....................................GT...T.............C.T...........A...T..A.............T......T...
6BBGc(6,1) ....
6BBGd(5,1) ....
6BBGf-1(7,1) C.....................
6BBGf-2(4,1) C.....................
6BBGf-3(1,1) C.....................
6BBGg-1(3,1) ....................................G.T..T..... ...
6BBGg-2(3,1) ....................................G.T..T..... ...
6BBGh(1,1) ....
6BBGi(1,1) ....
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
KC955130_BG8 CCATGTGCTGGGGCCATGAGCTATGTTGAGGCTTTGGAATGTGTTGGGGTTGTGGGATGTACTGGGGTCGTGGGATGTGTCAATCCTGGCTGATTCACGT
6BBGa-1(3,1) ..T.....A.........G.A.C.A. .G.A.C....G...........CC.........G..............C...........AT....C.T. .
6BBGa-2(1,1) ..T.....A.........G.A.C.A. .G.A.C....G...........CC.........G..............C...........AT....C.T. .
6BBGa-3(1,1)
6BBGb-1(17,2)
6BBGb-2(7,1)
6BBGb-3(3,1) .......T.A..TT....G.A.G.C...G...C...A...TAAG....T.CA..........CA.A..T..C..GAT.TGTT..TG.TAT..TGGG.TTG
6BBGc(6,1)
6BBGd(5,1)
6BBGf-1(7,1)
6BBGf-2(4,1)
6BBGf-3(1,1)
6BBGg-1(3,1) ..G.....
6BBGg-2(3,1) ..G.....
6BBGh(1,1)
6BBGi(1,1) .....T.T.....
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
KC955130_BG8 GGAAAAACCTTTCACAATCGGTTCCTTCCAGTTTGTTTAATTCCTTCTTGGGCCCAAAGTGGTCATTGGACTCCTCCCAGAAAAAAGGGTTTGGGGTCAG
6BBGa-1(3,1) TC.G..CT...G.C.............GA..CT....C.......... A.............C...T.ATAGA..G...AA.A..T..
6BBGa-2(1,1) TC.G..CT...G.C.............GA..CT....C..........A.............C...T.AAT....G..AA.A..T..
6BBGa-3(1,1)
6BBGb-1(17,2)
6BBGb-2(7,1)
6BBGb-3(3,1) TCG.TCCAGAC.GCTGCATATGAAAAA.TTA.CCCCAGTCG.T.CCT.CCCATTGTTTTAAT..C..CTTGGGAC.AA.TTGGTC.TT.GCCACTTC.CA
6BBGc(6,1)
6BBGd(5,1)
6BBGf-1(7,1)
6BBGf-2(4,1)
6BBGf-3(1,1) .
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1)
6BBGi(1,1) .....
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
KC955130_BG8 GGTGTGAGAGCTGATGGCATGGAAACGTGTCCCCTCTGACCATGCATTTCATTTGCTTCTATTTTGCAGAGAGAAAAGATGCAGAGTTGGGTAAGTCTCC
6BBGa-1(3,1)
6BBGa-2(1,1)
6BBGa-3(1,1)
6BBGb-1(17,2)
6BBGb-2(7,1)
6BBGb-3(3,1) .AACAA.A..A.TTG..GTCTAGGTATGCAAG.AG.CAGGGGATG.GAAG
6BBGc(6,1) .....
6BBGd(5,1)
6BBGf-1(7,1)
6BBGf-2(4,1)
6BBGf-3(1,1) TA..CA..CAGCC....G...AGG.G....T.....A....C.CA.....
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1)
6BBGi(1,1) .....C.....
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
KC955130_BG8 TTCCCTAAAGCGAGGGAATTCAGGGTGTCCCCATGGCATCAGCCGTGGAATTAGTAGCTGTCCTCTCTGACAATTCACTGCTCTGCTCTTTCCTTTCCAG
6BBGa-1(3,1) A..C.GGGT....G.G.....A...G..G..A..G....T.C.......C..G.....
6BBGa-2(1,1) A..C.GGGT....G.G.....A...G..G..A..G....T.C.......C..G.....
6BBGa-3(1,1)
6BBGb-1(17,2)
6BBGb-2(7,1)
6BBGb-3(3,1)
6BBGc(6,1)
6BBGd(5,1)
6BBGf-1(7,1)
6BBGf-2(4,1)
6BBGf-3(1,1)
6BBGg-1(3,1) .....G..CT.......C.......C..G.....
6BBGg-2(3,1) .....G..CT.......C.......C..G.....
6BBGh(1,1)
6BBGi(1,1)

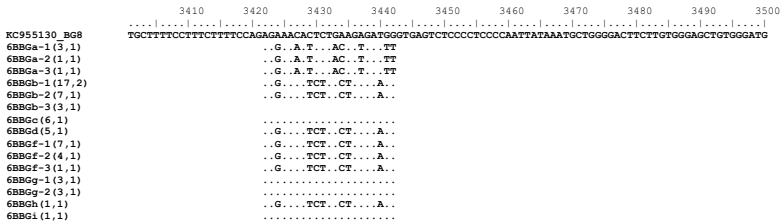

2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
KC955130_BG8 TGGAGAAAGCTGCAGCATTGGGTGAGTTATATTCCCCAAGCCAAAGTACTTTGGGTCTTCCCATTGGAAGTTATTTCCTCAGACCATCCTTTCTGTTGTG
6BBGa-1(3,1) ....CC.....T.....A.....
6BBGa-2(1,1) ....CC.....
6BBGa-3(1,1) ....CC.....
6BBGb-1(17,2)
6BBGb-2(7,1)
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) A..GA...A.....C...
6BBGf-1(7,1) A..G.....A.....C...
6BBGf-2(4,1) A..G.....A.....C...
6BBGf-3(1,1) A..G.....A.....C...
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) A..GA...A.....C...
6BBGi(1,1) .....
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
KC955130_BG8 TTTGCTTTGGC ATCATGTTAGTAAAATGCCTTCTTGGGACCAAAGTGGTCATTGGCCACTTCCCAGAAAAAA AGATTTGGGGGCAGGGTGTGGGAGCT
6BBGa-1(3,1) ..........TT.....G.G.....C...A..G.....T.....AA.....
6BBGa-2(1,1)
6BBGa-3(1,1)
6BBGb-1(17,2)
6BBGb-2(7,1)
6BBGb-3(3,1)
6BBGc(6,1)
6BBGd(5,1)
6BBGf-1(7,1)
6BBGf-2(4,1)
6BBGf-3(1,1)
6BBGg-1(3,1)
6BBGg-2(3,1)
6BBGh(1,1)
6BBGi(1,1)
2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
KC955130_BG8 GATGGCATGGAAATTTGTCCCCTCTGACCATGCTTTTCCTTTGCTTC TTTTTGCAGAGAGAAAAGATGCAGAGTTGGGTAAGTCTCCTTCCCCACAGTG
6BBGa-1(3,1) ....A.....G.....A.....C.T.....
6BBGa-2(1,1) .....
6BBGa-3(1,1) .....
6BBGb-1(17,2) .....G.....CAC...
6BBGb-2(7,1) .....G.....CAC...
6BBGb-3(3,1) ....T.....A.....A.....TTA.....G.....CAC...
6BBGc(6,1) .....
6BBGd(5,1)
6BBGf-1(7,1)
6BBGf-2(4,1)
6BBGf-3(1,1)
6BBGg-1(3,1) .....C.
6BBGg-2(3,1) .....
6BBGh(1,1)
6BBGi(1,1) .....
2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
KC955130_BG8 AGGGAATTCAGGGTTTCCCCATGGCGTTAGCCACGGGATGGGCAGCTGTCCTCTCTGACCATGCACTGCTCTGCTCTTTCTTTTCCAGCGGAACAAGCAG
6BBGa-1(3,1) T.....T..AT.
6BBGa-2(1,1) T.....T..AT.
6BBGa-3(1,1) T.....T..AT.
6BBGb-1(17,2) .....T.C.T.
6BBGb-2(7,1) .....T.C.T.
6BBGb-3(3,1) .....T.C.T.
6BBGc(6,1) .....
6BBGd(5,1) .....T.C.T.
6BBGf-1(7,1) .....T.C.T.
6BBGf-2(4,1) .....T.C.T.
6BBGf-3(1,1) .....T.C.T.
6BBGg-1(3,1) .....A..... .....
6BBGg-2(3,1) .....
6BBGh(1,1) .....T.C.T.
6BBGi(1,1) .....
2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
KC955130_BG8 CGCTATGGAGTGAGTCTCCCCCCTCCATTTTTATTATTTTTAAATGTTCAGCCTCCGGTAGCTGTGGGATGAGATGTTCCTCTCATCATACACTGACTCTG
6BBGa-1(3,1) A.A.....G
6BBGa-2(1,1) A.A.....G
6BBGa-3(1,1) A.A.....G
6BBGb-1(17,2) ..A...T.G
6BBGb-2(7,1) ..A...T.G
6BBGb-3(3,1) ..A...T.G
6BBGc(6,1) .....
6BBGd(5,1) ..A...T.G
6BBGf-1(7,1) ..A...T.G
6BBGf-2(4,1) ..A...T.G
6BBGf-3(1,1) ..A...T.G
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) ..A...T.G
6BBGi(1,1) ...A.....
2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
KC955130_BG8 CTTTTCCTTTGCAGAGCAAGAGATGCAATGTTGGGTGAGTCTCCCACCTGAAACCAAAGAGATTTGG GGTCTTCCCATGGGATCAGCCATGGGATGA
6BBGa-1(3,1) GTTT...T.C..A..GTC..A
6BBGa-2(1,1) GTTT...T.C..A..GTC..A
6BBGa-3(1,1) GTTT...T.C..A..GTC..A
6BBGb-1(17,2) GTGT.T.TAC.....ATC..A
6BBGb-2(7,1) GTGT.T.TAC.....ATC..A
6BBGb-3(3,1) GTGT.T.TAC.....ATC..A
6BBGc(6,1) .....
6BBGd(5,1) GTGT.T.TAC.....ATC..A
6BBGf-1(7,1) GTGT.T.T.C.....ATC..A
6BBGf-2(4,1) GTGT.T.T.C.....ATC..A
6BBGf-3(1,1) GTGT.T.T.C.....ATC..A
6BBGg-1(3,1) ....
6BBGg-2(3,1) .....
6BBGh(1,1) GTGT.T.TAC.....ATC..A
6BBGi(1,1) .....
2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
KC955130_BG8 TAACCTGAACCTTCTCATCGTGCGTTTCTTATTTGTTCCTTTTGCAGAGAAACACGTTCTAAAACTGGGTGAGTCCTCACTCCCAAATTATAAAGCAAAG
6BBGa-1(3,1) ..C..TTA.C.TC.....A
6BBGa-2(1,1) ..C..TTA.C.TC.....A
6BBGa-3(1,1) ..C..TTA.C.TC.....A
6BBGb-1(17,2) ...TC.TA.C.TC.....A
6BBGb-2(7,1) ...TC.TA.C.TC.....A
6BBGb-3(3,1) ...TC.TA.C.TC.....A
6BBGc(6,1) .....
6BBGd(5,1) ...TC.TA.C.TC.....A
6BBGf-1(7,1) ...TC.TA.C.TC.....A
6BBGf-2(4,1) ...TC.TA.C.TC.....A
6BBGf-3(1,1) ...TC.TA.C.TC.....A
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) ...TC.TA.C.TC.....A
6BBGi(1,1) .C.....
2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
KC955130_BG8 GGTTCTGCCTGTGTGAGCTGTGGGATCAGACGTTCCACTCATCATGCATTGCTTTTCTCTTTCTTTTTCAGAGGAAAAGACAGACGAAGTGGGTGAGTCT
6BBGa-1(3,1) .C.....TG.T.....
6BBGa-2(1,1) .C.....TG.T.....
6BBGa-3(1,1) .C.....TG.T.....
6BBGb-1(17,2) T.A..C.A.TG..AA..T...
6BBGb-2(7,1) T.A..C.A.TG..AA..T...
6BBGb-3(3,1) T.A..C.A.TG..AA..T...
6BBGc(6,1) .....
6BBGd(5,1) T.A..C.A.TG..AA..T...
6BBGf-1(7,1) T.A..C.A.TG..AA..T...
6BBGf-2(4,1) T.A..C.A.TG..AA..T...
6BBGf-3(1,1) T.A..C.A.TG..AA..T...
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) T.A..C.A.TG..AA..T...
6BBGi(1,1) .....
2810 2820 2830 2840 2850 2860 2870 2880 2890 2900
KC955130_BG8 ACATTCACTAAAGCAAAGAAATATGGGGTCTCCCATGGGATGACAAGCTGTCCCAAAAAATCATGTGGTGCTTTTTCTTGTCTTTTTATTTATTTATTTA
6BBGa-1(3,1)
6BBGa-2(1,1)
6BBGa-3(1,1)
6BBGb-1(17,2)
6BBGb-2(7,1)
6BBGb-3(3,1)
6BBGc(6,1)
6BBGd(5,1)
6BBGf-1(7,1)
6BBGf-2(4,1)
6BBGf-3(1,1)
6BBGg-1(3,1)
6BBGg-2(3,1)
6BBGh(1,1)
6BBGi(1,1)
2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
KC955130_BG8 TTTATTTATTTGCAGAGAATTGGAATTCAGTGCTGAGTAAGTTGCAGTCACTGAACTGAGGGAATGTGGGGGTCTTCCTAAGGGACTGCGTAGGGG
6BBGa-1(3,1) ..C....C....T..AT....
6BBGa-2(1,1) ..C....C....T..AT....
6BBGa-3(1,1) ..C....C....T..AT....
6BBGb-1(17,2) ...T.CA.....CT..G
6BBGb-2(7,1) ...T.CA.....CT..G
6BBGb-3(3,1) ...T.CA.....CT..G...T.....CA.....T..T....CC..C.GA....AATC..T...ACAAA
6BBGc(6,1) .....
6BBGd(5,1) ...T.CA.....CT..G
6BBGf-1(7,1) ...T.CA.....CT..G
6BBGf-2(4,1) ...T.CA.....CT..G
6BBGf-3(1,1) ...T.CA.....CT..G
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) ..T.CA.....CT..G
6BBGi(1,1) .....

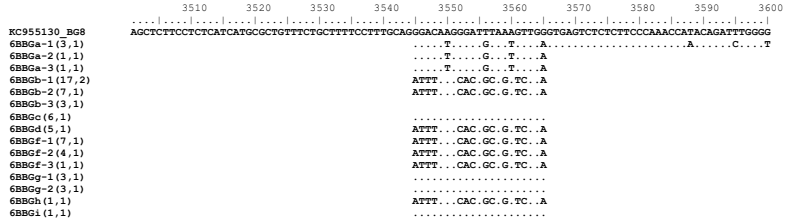

3010 3020 3030 3040 3050 3060 3070 3080 3090 3100
KC955130_BG8 AGAAGTTCCCATGCACTGCTTTTCTCTTTCTTTTCCAGAGAAAGACAGTGAAGA GATGGGTGAGTCTCTCCTCCCAAATTAAAAACGTTGGGGT
6BBGa-1(3,1) .....T..... C
6BBGa-2(1,1) .....T..... C
6BBGa-3(1,1) .....T..... C
6BBGb-1(17,2) .....CGGTA...GA. T.C..
6BBGb-2(7,1) .....CGGTA...GA. T.C..
6BBGb-3(3,1) AA TCCCCTC.GA....A..... .....CGGTA...GA. T.C..
6BBGc(6,1) ....
6BBGd(5,1) .....CGGTA...GA. T.C..
6BBGf-1(7,1) .....TGGTA...GA. T.C..
6BBGf-2(4,1) .....TGGTA...GA. T.C..
6BBGf-3(1,1) .....TGGTA...GA. T.C..
6BBGg-1(3,1) ....
6BBGg-2(3,1) ....
6BBGh(1,1) .....CGGTA...GA. T.C..
6BBGi(1,1) .A..... .....

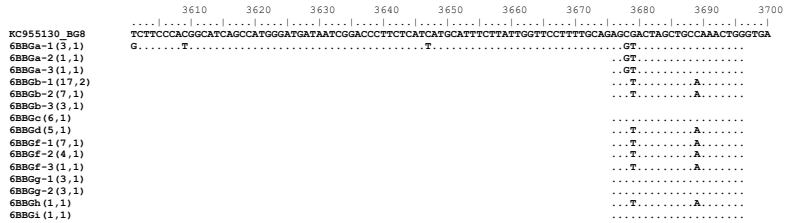

3110 3120 3130 3140 3150 3160 3170 3180 3190 3200
KC955130_BG8 TCCCATGTGGGAGCTGTGGGATGAGATGTTCCTCTCATCAACCATCTTTTTTACTTTTCCTTTGCAGGTTATGGCTTTGCAGAACTGAGTAAGTCTCCCT
6BBGa-1(3,1) ...C....G.......T....
6BBGa-2(1,1) ...C....G.......T....
6BBGa-3(1,1) ...C....G.......T....
6BBGb-1(17,2)
6BBGb-2(7,1)
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1)
6BBGf-1(7,1)
6BBGf-2(4,1)
6BBGf-3(1,1)
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1)
6BBGi(1,1) .....G...T.....

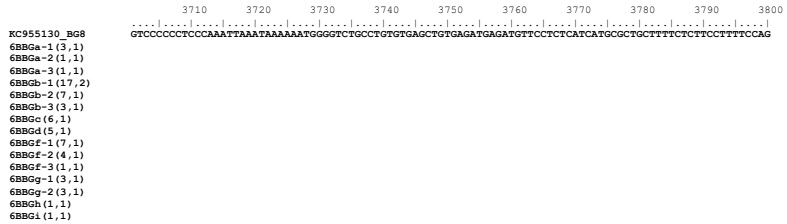

3210 3220 3230 3240 3250 3260 3270 3280 3290 3300
KC955130_BG8 CCCAACACGGAAGGGATTTGTGGTCTTCCCATGGGATCAGCCATGGGATGATCATCTGACCCCTCTCATCATGCATTTCGTATTTGTTTCTTTTGCAGAG
6BBGa-1(3,1) ..
6BBGa-2(1,1) ..
6BBGa-3(1,1) ..
6BBGb-1(17,2) ..
6BBGb-2(7,1) ..
6BBGb-3(3,1) ..
6BBGc(6,1) ..
6BBGd(5,1) ..
6BBGf-1(7,1) ..
6BBGf-2(4,1) ..
6BBGf-3(1,1) ..
6BBGg-1(3,1) ..
6BBGg-2(3,1) ..
6BBGh(1,1) ..
6BBGi(1,1) .....T.....A....A.....AA.....A...A.....C.....

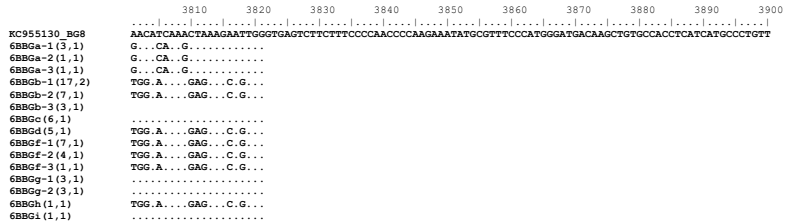

3310 3320 3330 3340 3350 3360 3370 3380 3390 3400
KC955130_BG8 AAACTGGCTGCAGAACTGGGTGAGTGCTGCCTCCCAAATTAAATAAAAAATGGGTCTGCCTGGGAGAGTGGTGGGATGGCATGTTCCTCTCACTGCGTGT
6BBGa-1(3,1) .....AG.....
6BBGa-2(1,1) .....AG.....
6BBGa-3(1,1) .....AG.....
6BBGb-1(17,2) G.....T....
6BBGb-2(7,1) G.....T....
6BBGb-3(3,1) G.....T.....
6BBGc(6,1) .....
6BBGd(5,1) G.....T....
6BBGf-1(7,1) G.....T....
6BBGf-2(4,1) G.....T....
6BBGf-3(1,1) G.....T....
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) G.....T....
6BBGi(1,1) .....

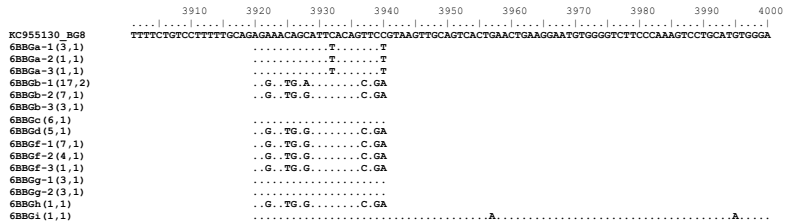

3410 3420 3430 3440 3450 3460 3470 3480 3490 3500
KC955130_BG8 TGCTTTTCCTTTCTTTTCCAGAGAAACACTTCGAAGAGATGGGTGAGTCTCCCCTCCCCAATTATAAATGCTGGGGACTTCTTGTGGGAGCTGTGGGATG
6BBGa-1(3,1) ..G..A.T...AC..T...TT
6BBGa-2(1,1) ..G..A.T...AC..T...TT
6BBGa-3(1,1) ..G..A.T...AC..T...TT
6BBGb-1(17,2) ..G....TCT..CT....A..
6BBGb-2(7,1) ..G....TCT..CT....A..
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) ..G....TCT..CT....A..
6BBGf-1(7,1) ..G....TCT..CT....A..
6BBGf-2(4,1) ..G....TCT..CT....A..
6BBGf-3(1,1) ..G....TCT..CT....A..
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) ..G....TCT..CT....A..
6BBGi(1,1) .....

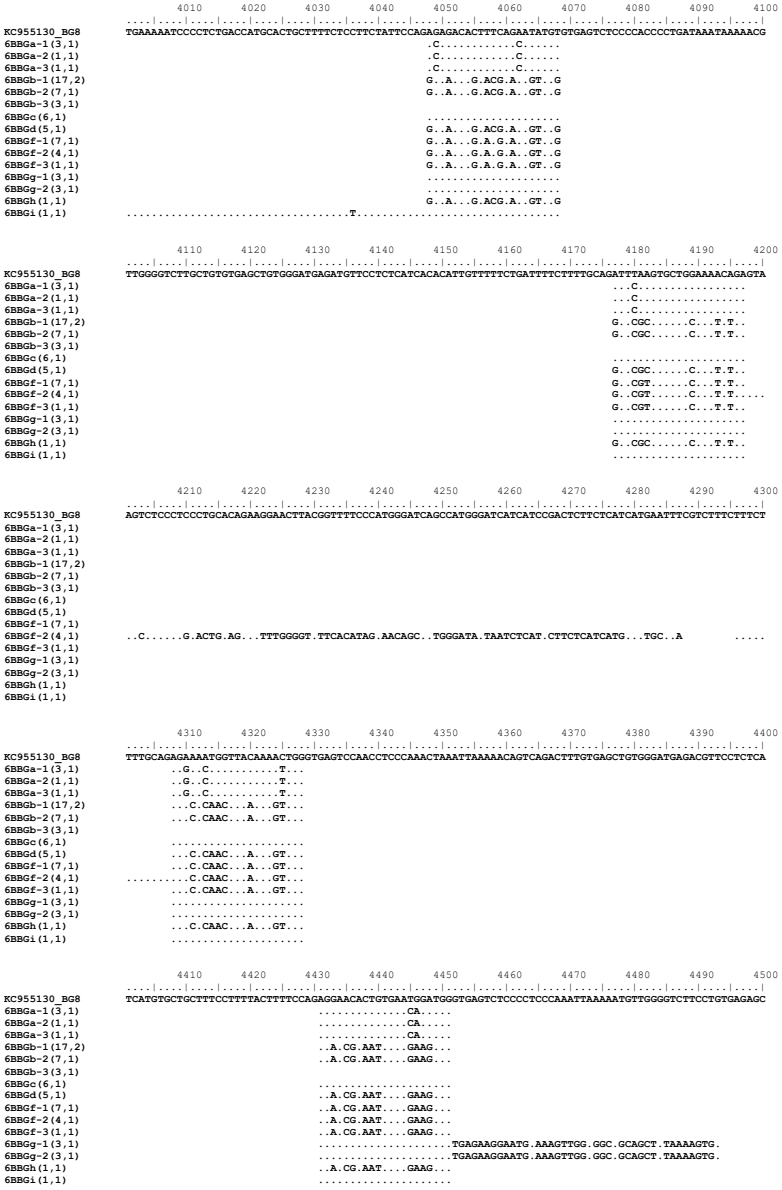

3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
KC955130_BG8 AGCTCTTCCTCTCATCATGCGCTGTTTCTGCTTTTCCTTTGCAGGGACAAGGGATTTAAAGTTGGGTGAGTCTCTTCCCAAACCATACAGATTTGGGG
6BBGa-1(3,1) .....T.....G...T....A.....A.....C....T
6BBGa-2(1,1) .....T.....G...T....A
6BBGa-3(1,1) .....T.....G...T....A
6BBGb-1(17,2) ATTT...CAC.GC.G.TC..A
6BBGb-2(7,1) ATTT...CAC.GC.G.TC..A
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) ATTT...CAC.GC.G.TC..A
6BBGf-1(7,1) ATTT...CAC.GC.G.TC..A
6BBGf-2(4,1) ATTT...CAC.GC.G.TC..A
6BBGf-3(1,1) ATTT...CAC.GC.G.TC..A
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) ATTT...CAC.GC.G.TC..A
6BBGi(1,1) .....

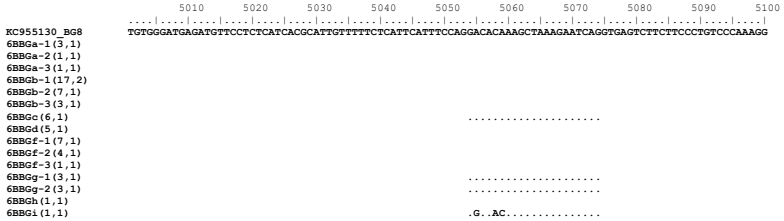

3610 3620 3630 3640 3650 3660 3670 3680 3690 3700
KC955130_BG8 TCTTCCCACGGCATCAGCCATGGGATGATAATCGGACCCTTCTCATCATGCATTTCTTATTGGTTCCTTTTGCAGAGCGACTAGCTGCCAAACTGGGTGA
6BBGa-1(3,1) G.....T.....T.....GT.....
6BBGa-2(1,1) .....GT.....
6BBGa-3(1,1) .....GT.....
6BBGb-1(17,2) .....T.....A.....
6BBGb-2(7,1) .....T.....A.....
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) .....T.....A.....
6BBGf-1(7,1) .....T.....A.....
6BBGf-2(4,1) .....T.....A.....
6BBGf-3(1,1) .....T.....A.....
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) .....T.....A.....
6BBGi(1,1) .....

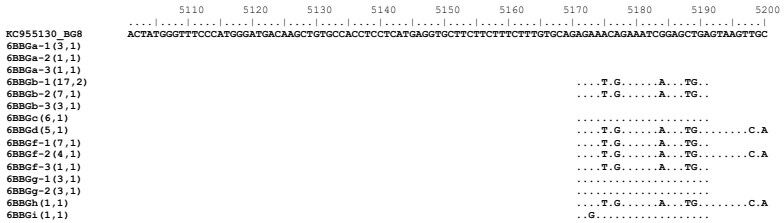

3710 3720 3730 3740 3750 3760 3770 3780 3790 3800
KC955130_BG8 GTCCCCCCTCCCAAATTAAATAAAAAATGGGGTCTGCCTGTGTGAGCTGTGAGATGAGATGTTCCTCTCATCATGCGCTGCTTTTCTCTTCCTTTTCCAG
6BBGa-1(3,1)
6BBGa-2(1,1)
6BBGa-3(1,1)
6BBGb-1(17,2)
6BBGb-2(7,1)
6BBGb-3(3,1)
6BBGc(6,1)
6BBGd(5,1)
6BBGf-1(7,1)
6BBGf-2(4,1)
6BBGf-3(1,1)
6BBGg-1(3,1)
6BBGg-2(3,1)
6BBGh(1,1)
6BBGi(1,1)

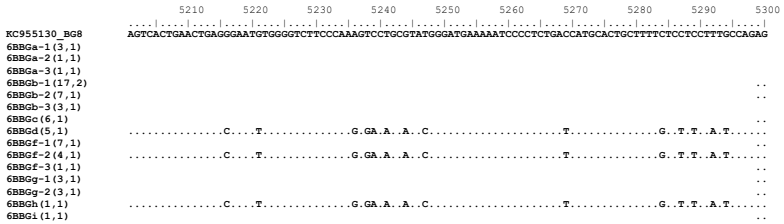

3810 3820 3830 3840 3850 3860 3870 3880 3890 3900
KC955130_BG8 AACATCAAACTAAAGAATTGGGTGAGTCTTCTTTCCCCAACCCCAAGAAATATGCGTTTCCCATGGGATGACAAGCTGTGCCACCTCATCATGCCCTGTT
6BBGa-1(3,1) G...CA..G.....
6BBGa-2(1,1) G...CA..G.....
6BBGa-3(1,1) G...CA..G.....
6BBGb-1(17,2) TGG.A....GAG...C.G...
6BBGb-2(7,1) TGG.A....GAG...C.G...
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) TGG.A....GAG...C.G...
6BBGf-1(7,1) TGG.A....GAG...C.G...
6BBGf-2(4,1) TGG.A....GAG...C.G...
6BBGf-3(1,1) TGG.A....GAG...C.G...
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) TGG.A....GAG...C.G...
6BBGi(1,1) .....

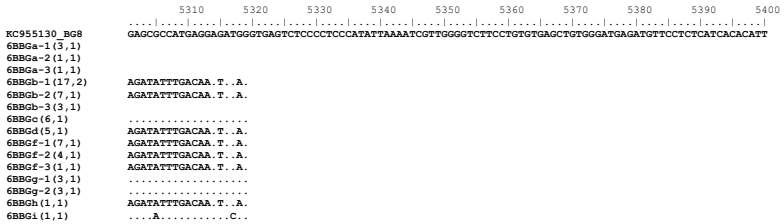

3910 3920 3930 3940 3950 3960 3970 3980 3990 4000
KC955130_BG8 TTTTCTGTCCTTTTTGCAGAGAAACAGCATTCACAGTTCCGTAAGTTGCAGTCACTGAACTGAAGGAATGTGGGGTCTTCCCAAAGTCCTGCATGTGGGA
6BBGa-1(3,1) .....T.....T
6BBGa-2(1,1) .....T.....T
6BBGa-3(1,1) .....T.....T
6BBGb-1(17,2) ..G..TG.A.....C.GA
6BBGb-2(7,1) ..G..TG.G.....C.GA
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) ..G..TG.G.....C.GA
6BBGf-1(7,1) ..G..TG.G.....C.GA
6BBGf-2(4,1) ..G..TG.G.....C.GA
6BBGf-3(1,1) ..G..TG.G.....C.GA
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) ..G..TG.G.....C.GA
6BBGi(1,1) .....A.....A....

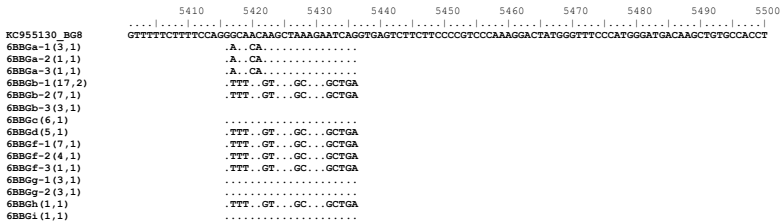

4010 4020 4030 4040 4050 4060 4070 4080 4090 4100
KC955130_BG8 TGAAAAATCCCCTCTGACCATGCACTGCTTTTCTCCTTCTATTCCAGAGAGACACTTTCAGAATATGTGTGAGTCTCCCCCACCCCTGATAAATAAAAACG
6BBGa-1(3,1) .C.....C.....
6BBGa-2(1,1) .C.....C.....
6BBGa-3(1,1) .C.....C.....
6BBGb-1(17,2) G..A...G.ACG.A..GT..G
6BBGb-2(7,1) G..A...G.ACG.A..GT..G
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) G..A...G.ACG.A..GT..G
6BBGf-1(7,1) G..A...G.A.G.A..GT..G
6BBGf-2(4,1) G..A...G.A.G.A..GT..G
6BBGf-3(1,1) G..A...G.A.G.A..GT..G
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) G..A...G.ACG.A..GT..G
6BBGi(1,1) .....T.....
4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
KC955130_BG8 TTGGGGTCTTGCTGTGTGAGCTGTGGGATGAGATGTTCCTCTCATCACACATTGTTTTTCTGATTTTCTTTTGCAGATTTAAGTGCTGGAAAACAGAGTA
6BBGa-1(3,1) ...C.....
6BBGa-2(1,1) ...C.....
6BBGa-3(1,1) ...C.....
6BBGb-1(17,2) G..CGC.....C...T.T..
6BBGb-2(7,1) G..CGC.....C...T.T..
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) G..CGC.....C...T.T..
6BBGf-1(7,1) G..CGT.....C...T.T..
6BBGf-2(4,1) G..CGT.....C...T.T.....
6BBGf-3(1,1) G..CGT.....C...T.T..
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) G..CGC.....C...T.T..
6BBGi(1,1) .....
4210 4220 4230 4240 4250 4260 4270 4280 4290 4300
KC955130_BG8 AGTCTCCCTCCCTGCACAGAAGGAACTTACGGTTTTCCCATGGGATCAGCCATGGGATCATCATCCGACTCTTCTCATCATGAATTTCGTCTTTCTTTCT
6BBGa-1(3,1)
6BBGa-2(1,1)
6BBGa-3(1,1)
6BBGb-1(17,2)
6BBGb-2(7,1)
6BBGb-3(3,1)
6BBGc(6,1)
6BBGd(5,1)
6BBGf-1(7,1)
6BBGf-2(4,1) ..C.....G.ACTG.AG...TTTGGGGT.TTCACATAG.AACAGC..TGGGATA.TAATCTCAT.CTTCTCATCATG...TGC..A .....
6BBGf-3(1,1)
6BBGg-1(3,1)
6BBGg-2(3,1)
6BBGh(1,1)
6BBGi(1,1)
4310 4320 4330 4340 4350 4360 4370 4380 4390 4400
KC955130_BG8 TTTGCAGAGAAAATGGTTACAAAACTGGGTGAGTCCAACCTCCCAAACTAAATTAAAAACAGTCAGACTTTGTGAGCTGTGGGATGAGACGTTCCTCTCA
6BBGa-1(3,1) ..G..C.....T...
6BBGa-2(1,1) ..G..C.....T...
6BBGa-3(1,1) ..G..C.....T...
6BBGb-1(17,2) ...C.CAAC...A...GT...
6BBGb-2(7,1) ...C.CAAC...A...GT...
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) ...C.CAAC...A...GT...
6BBGf-1(7,1) ...C.CAAC...A...GT...
6BBGf-2(4,1) .....C.CAAC...A...GT...
6BBGf-3(1,1) ...C.CAAC...A...GT...
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) ...C.CAAC...A...GT...
6BBGi(1,1) .....
4410 4420 4430 4440 4450 4460 4470 4480 4490 4500
KC955130_BG8 TCATGTGCTGCTTTCCTTTTACTTTTCCAGAGGAACACTGTGAATGGATGGGTGAGTCTCCCCTCCCAAATTAAAAATGTTGGGGTCTTCCTGTGAGAGC
6BBGa-1(3,1) .....CA.....
6BBGa-2(1,1) .....CA.....
6BBGa-3(1,1) .....CA.....
6BBGb-1(17,2) ..A.CG.AAT....GAAG...
6BBGb-2(7,1) ..A.CG.AAT....GAAG...
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) ..A.CG.AAT....GAAG...
6BBGf-1(7,1) ..A.CG.AAT....GAAG...
6BBGf-2(4,1) ..A.CG.AAT....GAAG...
6BBGf-3(1,1) ..A.CG.AAT....GAAG...
6BBGg-1(3,1) .....TGAGAAGGAATG.AAAGTTGG.GGC.GCAGCT.TAAAAGTG.
6BBGg-2(3,1) .....TGAGAAGGAATG.AAAGTTGG.GGC.GCAGCT.TAAAAGTG.
6BBGh(1,1) ..A.CG.AAT....GAAG...
6BBGi(1,1) .....
4510 4520 4530 4540 4550 4560 4570 4580 4590 4600
KC955130_BG8 TGTGGGATGAGCTGTTCCTCTCATCGTGCACTGTTTCTGCTTTTCCTTTGCAGTGAGAAGGAATGTAAAGTTGGGTGAGTCTTCTCCCCAACCAAAGAG
6BBGa-1(3,1) A.....C.....
6BBGa-2(1,1) A.....C.....
6BBGa-3(1,1) A.....C.....
6BBGb-1(17,2) A..A.CACCT.AA....A.T.
6BBGb-2(7,1) A..A.CACCT.AA....A.T.
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) A..A.CACCT.AA....A.T.
6BBGf-1(7,1) A..A.CACCT.AA....A.T.
6BBGf-2(4,1) A..A.CACCT.AA....A.T.
6BBGf-3(1,1) A..A.CACCT.AA....A.T.
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) A..A.CACCT.AA....A.T.
6BBGi(1,1) .....
4610 4620 4630 4640 4650 4660 4670 4680 4690 4700
KC955130_BG8 ATTCGGAGTCTTCCATGGGATCAGCCATGGGATGATAACATGAACCTCATCACGTGTTTCTTATTTGTTCCTTTTGCAGAGGCAGCAGCTGTAAAAGTGG
6BBGa-1(3,1) .....C...
6BBGa-2(1,1) .....C...
6BBGa-3(1,1) .....C...
6BBGb-1(17,2) GTAT.CGT...CCT..TC..A
6BBGb-2(7,1) GTAT.CGT...CCT..TC..A
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) GTAT.CGT...CCT..TC..A
6BBGf-1(7,1) GTAT.CGT...CCT..TC..A
6BBGf-2(4,1) GTAT.CGT...CCT..TC..A
6BBGf-3(1,1) GTAT.CGT...CCT..TC..A
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) GTAT.CGT...CCT..TC..A
6BBGi(1,1) ..AT.C.....
4710 4720 4730 4740 4750 4760 4770 4780 4790 4800
KC955130_BG8 GTGAGTCTCCCCTCCCAAATTAAAAATGTTGGCGTCATCCTGTGAGAGCTGTGGGATGAGCTGTTCCTCTCATCGTGCACTGTTTCTGCTTTTCCTTTGC
6BBGa-1(3,1)
6BBGa-2(1,1)
6BBGa-3(1,1)
6BBGb-1(17,2)
6BBGb-2(7,1)
6BBGb-3(3,1)
6BBGc(6,1)
6BBGd(5,1)
6BBGf-1(7,1)
6BBGf-2(4,1)
6BBGf-3(1,1)
6BBGg-1(3,1)
6BBGg-2(3,1)
6BBGh(1,1)
6BBGi(1,1)
4810 4820 4830 4840 4850 4860 4870 4880 4890 4900
KC955130_BG8 AGTGAGAAGGAATGTAAAGTTGGGTGAGTCTTCTTCCCCAACCAAAGAGATGTGGGGTCTTCCATGGGATCAGCCATGGGATGATAAGCTGAACCTTATC
6BBGa-1(3,1)
6BBGa-2(1,1)
6BBGa-3(1,1)
6BBGb-1(17,2) G.CT.CAC.TG.C.G.AC...
6BBGb-2(7,1) G.CT.CAC.TG.C.G.AC...
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) G.CT.CAC.TG.C.G.AC...
6BBGf-1(7,1) G.CT.CAC.TG.C.G.AC...
6BBGf-2(4,1) G.CT.CAC.TG.C.G.AC...
6BBGf-3(1,1) G.CT.CAC.TG.C.G.AC...
6BBGg-1(3,1)
6BBGg-2(3,1)
6BBGh(1,1) G.CT.CAC.TG.C.G.AC...
6BBGi(1,1)
4910 4920 4930 4940 4950 4960 4970 4980 4990 5000
KC955130_BG8 ACGTGTTTCTTATTTGTTCCTTTTGCAGAGGCAGCAGCTGTAAAAGTGGGTGAGTCCTCCCTCCCAAATCAAATACAAAAGGGGATCTGCCTGTGTGAGC
6BBGa-1(3,1)
6BBGa-2(1,1)
6BBGa-3(1,1)
6BBGb-1(17,2) T..ATCA..T..A.GC.....
6BBGb-2(7,1) T..ATCA..T..A.GC.....
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) T..ATCA..T..A.GC.....
6BBGf-1(7,1) T..ATCA..T..A.GC.....
6BBGf-2(4,1) T..ATCA..T..A.GC.....
6BBGf-3(1,1) T..ATCA..T..A.GC.....
6BBGg-1(3,1)
6BBGg-2(3,1)
6BBGh(1,1) T..ATCA..T..A.GC.....
6BBGi(1,1)

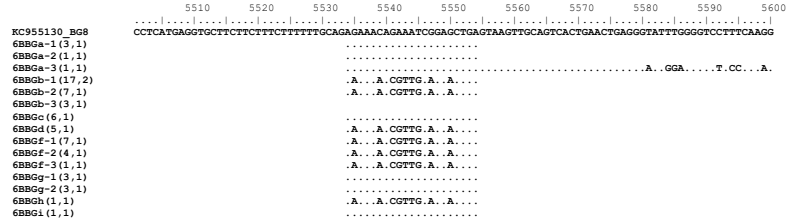

5010 5020 5030 5040 5050 5060 5070 5080 5090 5100
KC955130_BG8 TGTGGGATGAGATGTTCCTCTCATCACGCATTGTTTTTCTCATTCATTTTCCAGGACACAAAGCTAAAGAATCAGGTGAGTCTTCTTCCCTGTCCCAAAGG
6BBGa-1(3,1)
6BBGa-2(1,1)
6BBGa-3(1,1)
6BBGb-1(17,2)
6BBGb-2(7,1)
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1)
6BBGf-1(7,1)
6BBGf-2(4,1)
6BBGf-3(1,1)
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1)
6BBGi(1,1) .G..AC.....

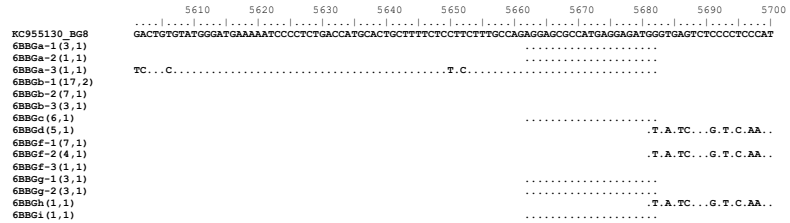

5110 5120 5130 5140 5150 5160 5170 5180 5190 5200
KC955130_BG8 ACTATGGGTTTCCCATGGGATGACAAGCTGTGCCACCTCCTCATGAGGTGCTTCTTCTTTCTTTGTGCAGAGAAACAGAAATCGGAGCTGAGTAAGTTGC
6BBGa-1(3,1)
6BBGa-2(1,1)
6BBGa-3(1,1)
6BBGb-1(17,2) ....T.G.....A...TG..
6BBGb-2(7,1) ....T.G.....A...TG..
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) ....T.G.....A...TG.....C.A
6BBGf-1(7,1) ....T.G.....A...TG..
6BBGf-2(4,1) ....T.G.....A...TG.....C.A
6BBGf-3(1,1) ....T.G.....A...TG..
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) ....T.G.....A...TG.....C.A
6BBGi(1,1) ..G.....

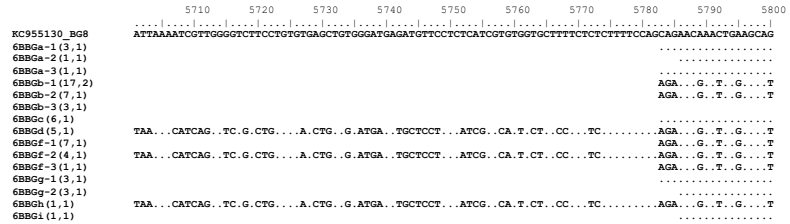

5210 5220 5230 5240 5250 5260 5270 5280 5290 5300
KC955130_BG8 AGTCACTGAACTGAGGGAATGTGGGGTCTTCCCAAAGTCCTGCGTATGGGATGAAAAATCCCCTCTGACCATGCACTGCTTTTCTCCTCCTTTGCCAGAG
6BBGa-1(3,1)
6BBGa-2(1,1)
6BBGa-3(1,1)
6BBGb-1(17,2) ..
6BBGb-2(7,1) ..
6BBGb-3(3,1)
6BBGc(6,1) ..
6BBGd(5,1) .....C....T.....G.GA.A..A..C.....T.....G..T.T..A.T.....
6BBGf-1(7,1) ..
6BBGf-2(4,1) .....C....T.....G.GA.A..A..C.....T.....G..T.T..A.T.....
6BBGf-3(1,1) ..
6BBGg-1(3,1) ..
6BBGg-2(3,1) ..
6BBGh(1,1) .....C....T.....G.GA.A..A..C.....T.....G..T.T..A.T.....
6BBGi(1,1) ..

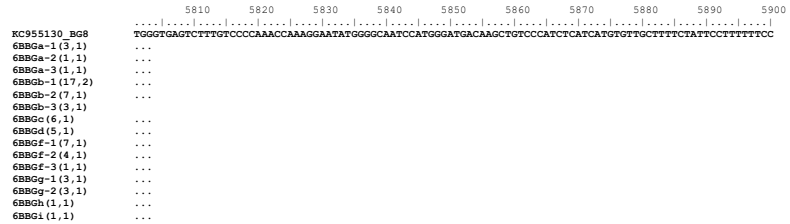

5310 5320 5330 5340 5350 5360 5370 5380 5390 5400
KC955130_BG8 GAGCGCCATGAGGAGATGGGTGAGTCTCCCCTCCCATATTAAAATCGTTGGGGTCTTCCTGTGTGAGCTGTGGGATGAGATGTTCCTCTCATCACACATT
6BBGa-1(3,1)
6BBGa-2(1,1)
6BBGa-3(1,1)
6BBGb-1(17,2) AGATATTTGACAA.T..A.
6BBGb-2(7,1) AGATATTTGACAA.T..A.
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) AGATATTTGACAA.T..A.
6BBGf-1(7,1) AGATATTTGACAA.T..A.
6BBGf-2(4,1) AGATATTTGACAA.T..A.
6BBGf-3(1,1) AGATATTTGACAA.T..A.
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) AGATATTTGACAA.T..A.
6BBGi(1,1) ....A.....C..

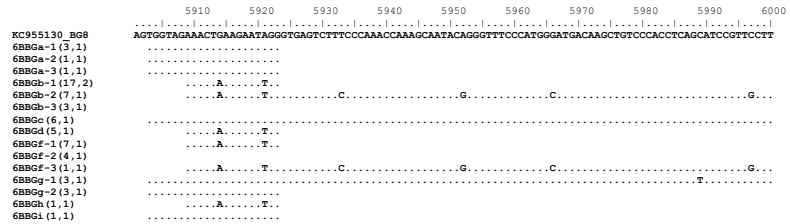

5410 5420 5430 5440 5450 5460 5470 5480 5490 5500
KC955130_BG8 GTTTTTCTTTTCCAGGGCAACAAGCTAAAGAATCAGGTGAGTCTTCTTCCCCGTCCCAAAGGACTATGGGTTTCCCATGGGATGACAAGCTGTGCCACCT
6BBGa-1(3,1) .A..CA.....
6BBGa-2(1,1) .A..CA.....
6BBGa-3(1,1) .A..CA.....
6BBGb-1(17,2) .TTT..GT...GC...GCTGA
6BBGb-2(7,1) .TTT..GT...GC...GCTGA
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) .TTT..GT...GC...GCTGA
6BBGf-1(7,1) .TTT..GT...GC...GCTGA
6BBGf-2(4,1) .TTT..GT...GC...GCTGA
6BBGf-3(1,1) .TTT..GT...GC...GCTGA
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) .TTT..GT...GC...GCTGA
6BBGi(1,1) .....

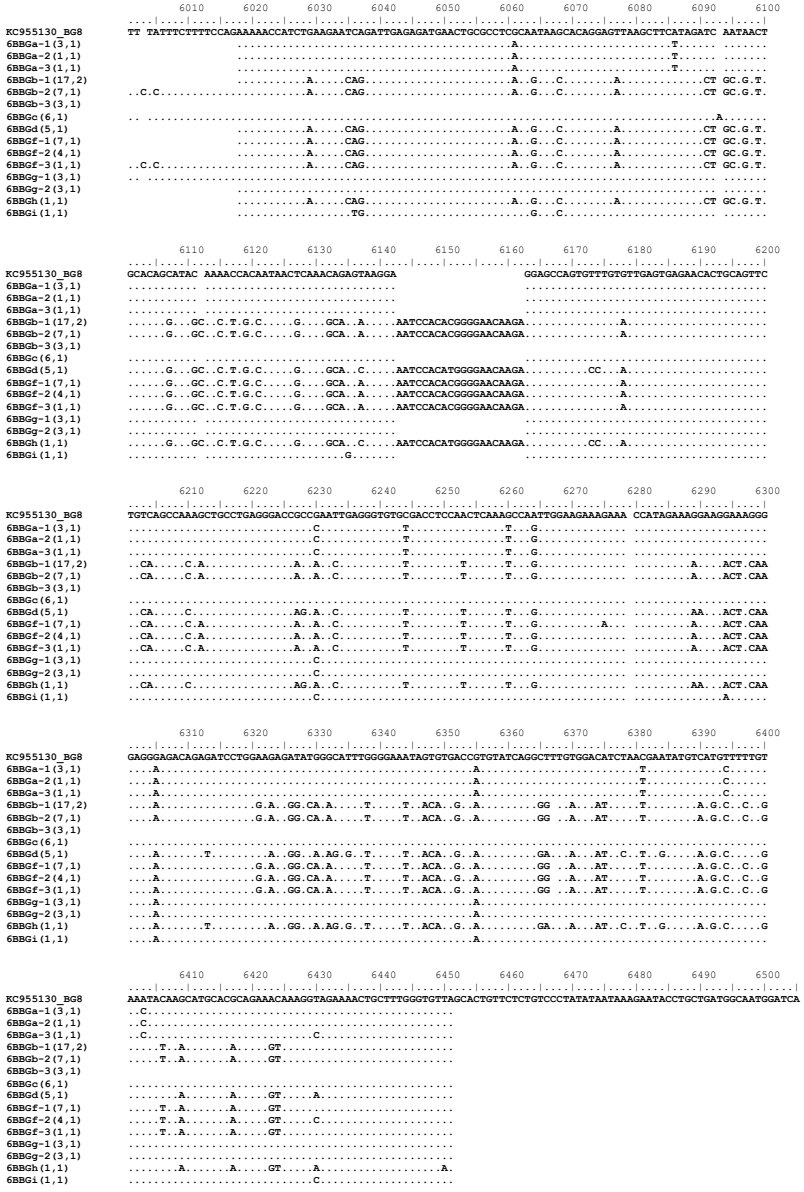

5510 5520 5530 5540 5550 5560 5570 5580 5590 5600
KC955130_BG8 CCTCATGAGGTGCTTCTTCTTTCTTTTTGCAGAGAAACAGAAATCGGAGCTGAGTAAGTTGCAGTCACTGAACTGAGGGTATTTGGGGTCCTTTCAAGG
6BBGa-1(3,1) .....
6BBGa-2(1,1) .....
6BBGa-3(1,1) .....A..GGA.....T.CC...A.
6BBGb-1(17,2) .A...A.CGTTG.A..A....
6BBGb-2(7,1) .A...A.CGTTG.A..A....
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) .A...A.CGTTG.A..A....
6BBGf-1(7,1) .A...A.CGTTG.A..A....
6BBGf-2(4,1) .A...A.CGTTG.A..A....
6BBGf-3(1,1) .A...A.CGTTG.A..A....
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) .A...A.CGTTG.A..A....
6BBGi(1,1) .....


5610 5620 5630 5640 5650 5660 5670 5680 5690 5700
KC955130_BG8 GACTGTGTATGGGATGAAAAATCCCCTCTGACCATGCACTGCTTTTCTCCTTCTTTGCCAGAGGAGCGCCATGAGGAGATGGGTGAGTCTCCCCTCCCAT
6BBGa-1(3,1) .....
6BBGa-2(1,1) .....
6BBGa-3(1,1) TC...C.....T.C.....
6BBGb-1(17,2)
6BBGb-2(7,1)
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) .T.A.TC...G.T.C.AA..
6BBGf-1(7,1)
6BBGf-2(4,1) .T.A.TC...G.T.C.AA..
6BBGf-3(1,1)
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) .T.A.TC...G.T.C.AA..
6BBGi(1,1) .....


5710 5720 5730 5740 5750 5760 5770 5780 5790 5800
KC955130_BG8 ATTAAAATCGTTGGGGTCTTCCTGTGTGAGCTGTGGGATGAGATGTTCCTCTCATCGTGTGGTGCTTTTCTCTCTTTTCCAGCAGAACAAACTGAAGCAG
6BBGa-1(3,1) .....
6BBGa-2(1,1) .....
6BBGa-3(1,1) .....
6BBGb-1(17,2) AGA...G..T..G....T
6BBGb-2(7,1) AGA...G..T..G....T
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) TAA...CATCAG..TC.G.CTG....A.CTG..G.ATGA..TGCTCCT...ATCG..CA.T.CT..CC...TC.....AGA...G..T..G....T
6BBGf-1(7,1) AGA...G..T..G....T
6BBGf-2(4,1) TAA...CATCAG..TC.G.CTG....A.CTG..G.ATGA..TGCTCCT...ATCG..CA.T.CT..CC...TC.....AGA...G..T..G....T
6BBGf-3(1,1) AGA...G..T..G....T
6BBGg-1(3,1) .....
6BBGg-2(3,1) .....
6BBGh(1,1) TAA...CATCAG..TC.G.CTG....A.CTG..G.ATGA..TGCTCCT...ATCG..CA.T.CT..CC...TC.....AGA...G..T..G....T
6BBGi(1,1) .....

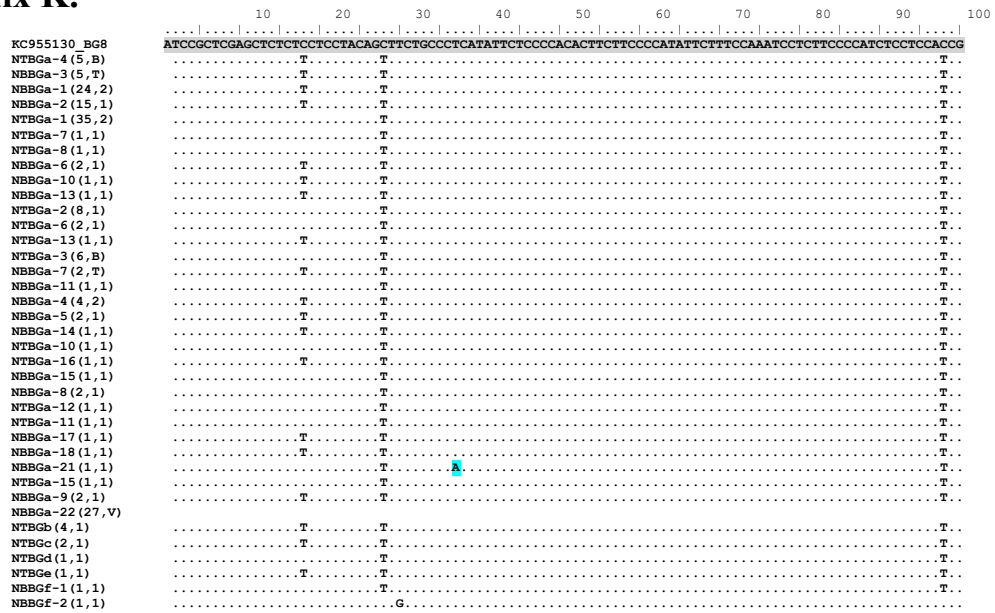

5810 5820 5830 5840 5850 5860 5870 5880 5890 5900
KC955130_BG8 TGGGTGAGTCTTTGTCCCCAAACCAAAGGAATATGGGGCAATCCATGGGATGACAAGCTGTCCCATCTCATCATGTGTTGCTTTTCTATTCCTTTTTTCC
6BBGa-1(3,1) ...
6BBGa-2(1,1) ...
6BBGa-3(1,1) ...
6BBGb-1(17,2) ...
6BBGb-2(7,1) ...
6BBGb-3(3,1)
6BBGc(6,1) ...
6BBGd(5,1) ...
6BBGf-1(7,1) ...
6BBGf-2(4,1) ...
6BBGf-3(1,1) ...
6BBGg-1(3,1) ...
6BBGg-2(3,1) ...
6BBGh(1,1) ...
6BBGi(1,1) ...

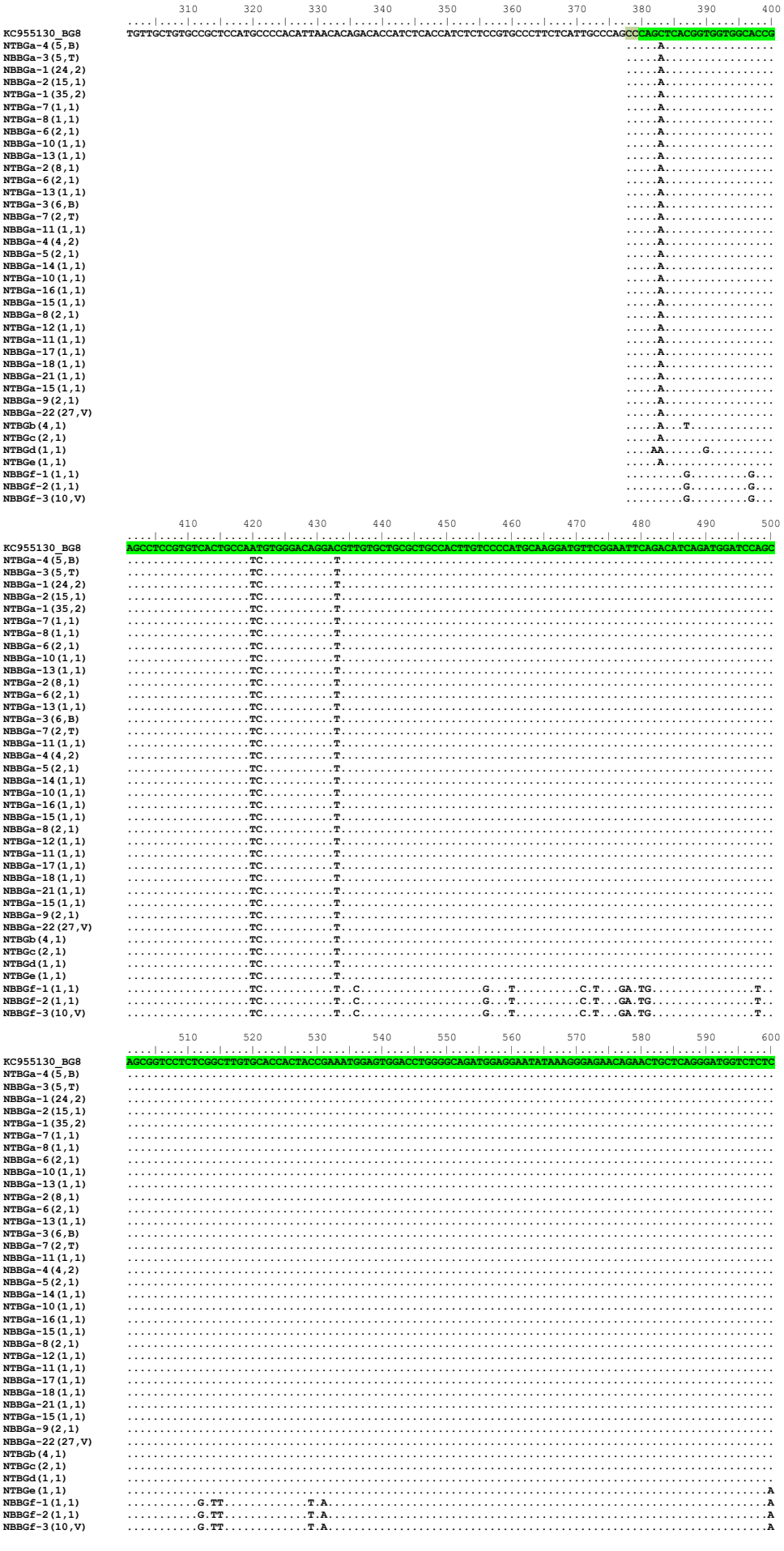

5910 5920 5930 5940 5950 5960 5970 5980 5990 6000
KC955130_BG8 AGTGGTAGAAACTGAAGAATAGGGTGAGTCTTTCCCAAACCAAAGCAATACAGGGTTTCCCATGGGATGACAAGCTGTCCCACCTCAGCATCCGTTCCTT
6BBGa-1(3,1) .....
6BBGa-2(1,1) .....
6BBGa-3(1,1) .....
6BBGb-1(17,2) .....A.....T..
6BBGb-2(7,1) .....A.....T.....C.....G.....C.....G...
6BBGb-3(3,1)
6BBGc(6,1) .....
6BBGd(5,1) .....A.....T..
6BBGf-1(7,1) .....A.....T..
6BBGf-2(4,1)
6BBGf-3(1,1) .....A.....T.....C.....G.....C.....G...
6BBGg-1(3,1) .....T.....
6BBGg-2(3,1) .....
6BBGh(1,1) .....A.....T..
6BBGi(1,1) .....

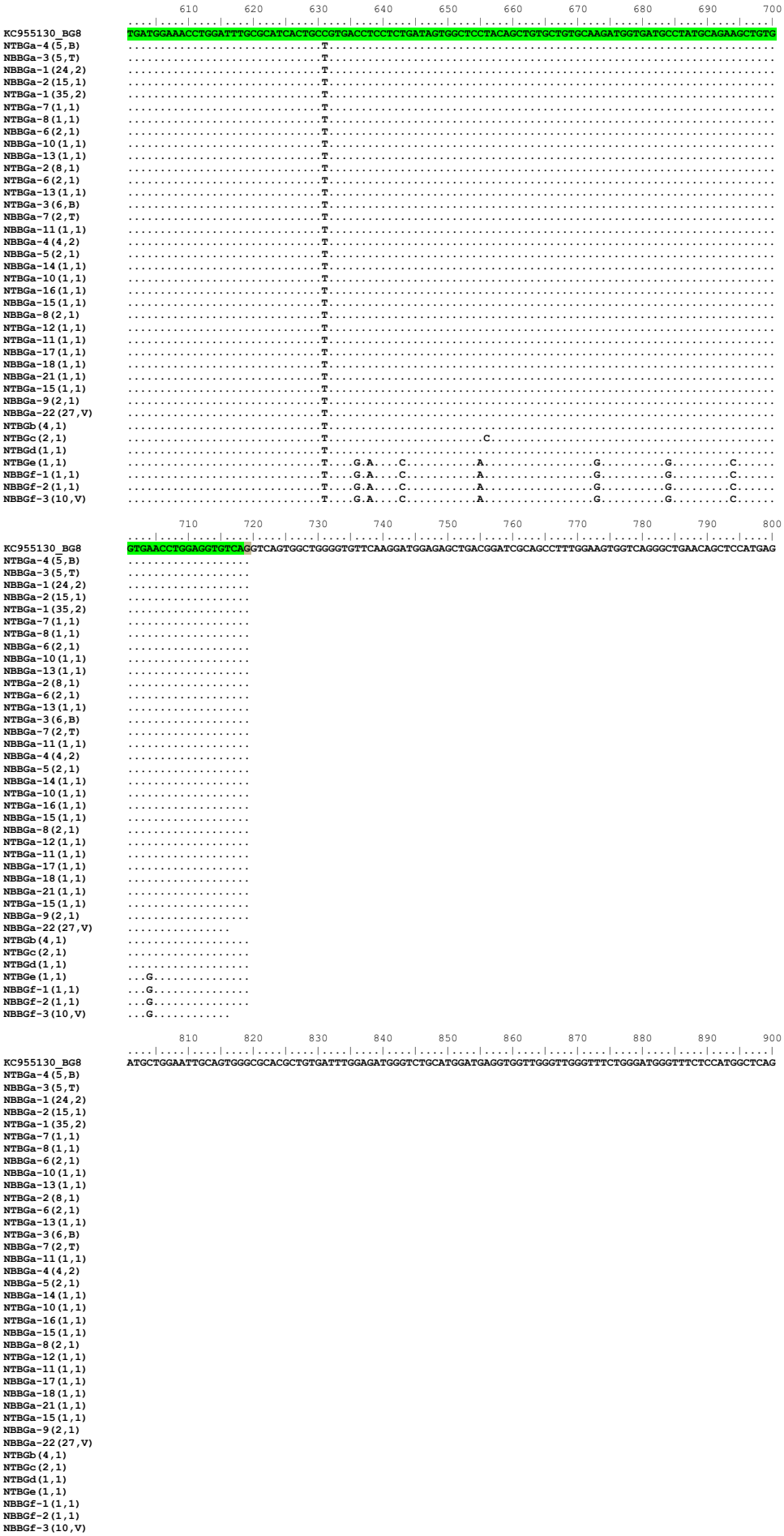


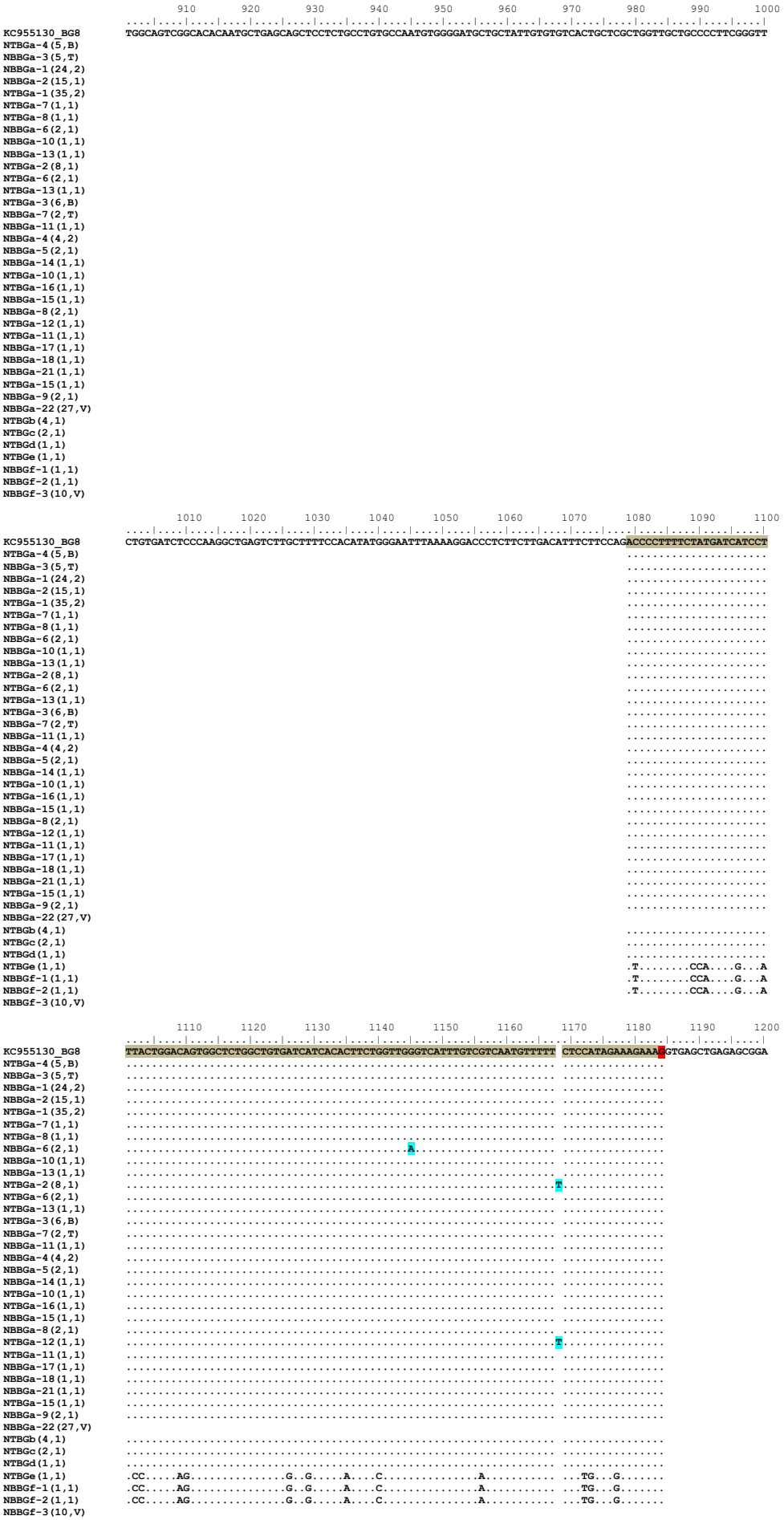
Appendix J. Alignments of all the cDNA sequences for the 18 genes found in B cells from four chicken lines against the BG8 gene (including introns) of B12 haplotype. Names of the gene follow the convention: GenBank accession number of B12 BG region genomic sequence, a dash, "BG" and the number of the gene locus for the B12 haplotype. Names of the transcripts follow the convention: abbreviated line name, "B" for B cells, "BG", a letter representing the exon 2 sequence, a dash and then a number representing the alternative splicing variant with "1" being the most frequently detected clone (and "2" the second most frequently detected clone, and so forth). Numbers in parenthesis indicate the number of times the depicted sequence was found out of the number of times the gene (based on the exon 2 sequence) was found, followed by the number of independent PCRs in which the sequence was identified (1, found in one PCR; 2 found in 2 PCRs; T, found in one PCR described in this paper and one using T cell cDNA). Letters indicate nucleotides, and dots indicate identity with BG8 sequence. (J1), the alignments between cDNA sequences found in line N (B21) and BG8 (B12); (J2) to (J4), between line P2a (B19), line 15I (B15) and line 6_1 (B2) to BG8 (B12), respectively.

Appendix K.

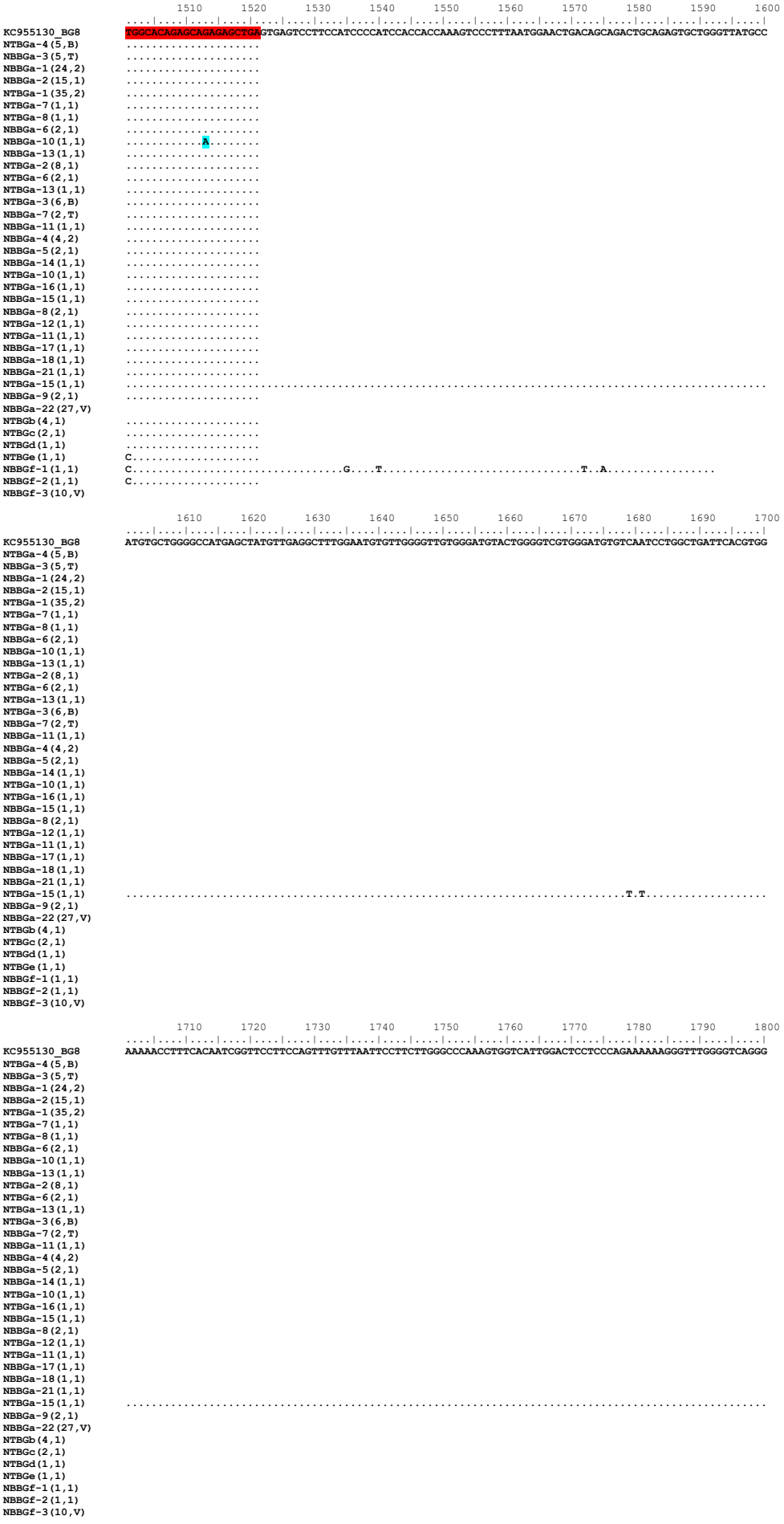

10 20 30 40 50 60 70 80 90 100
KC955130_BG8 ATCCGCTCGAGCTCTCTCCTCCTACAGCTTCTGCCCTCATATTCTCCCCACACTTCTTCCCCATATTCTTTCCAAATCCTCTTCCCCATCTCCTCCACCG
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
110 120 130 140 150 160 170 180 190 200
KC955130_BG8 TCTCCTTCTCAGAGTCCTTCCTCTCTCTCCCTAAATTCTTCCCCCCTCCTCTTCTCCAGCACAGATGGCCTTCACATCGGGCTGCAACCACCCCAGTTTC
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
210 220 230 240 250 260 270 280 290 300
KC955130_BG8 GCCCTCCCCTGGAGGACCCTCCTGCCTTATCTCGTGGCTCTGCACTTCCTCCAGCCGGGATCAGGTAGGGGTCCTGTGGGGCTGCTGTGCCTGGCACAGG
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)

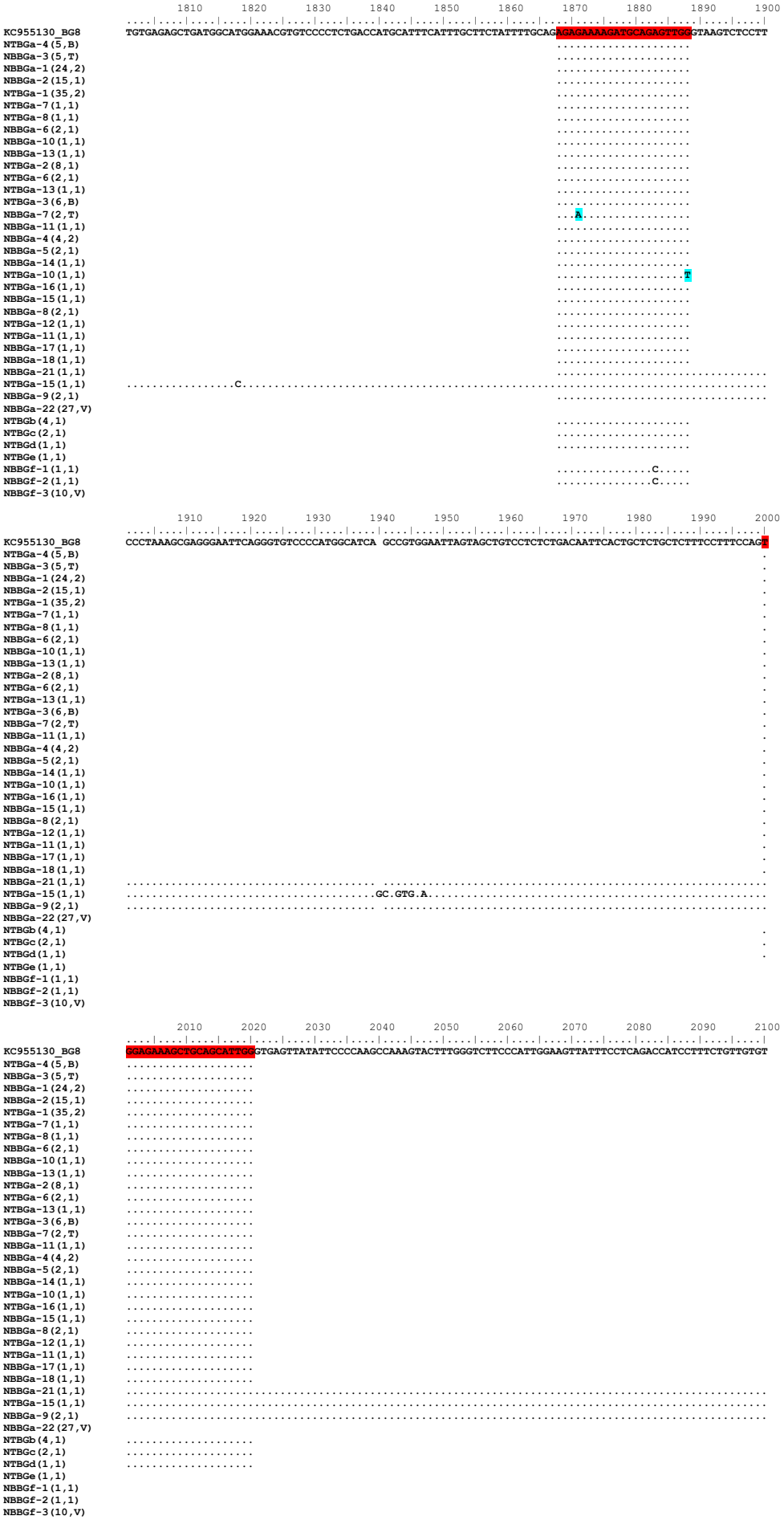

310 320 330 340 350 360 370 380 390 400
KC955130_BG8 TGTTGCTGTGCCGCTCCATGCCCCACATTAACACAGACACCATCTCACCATCTCTCCGTGCCCTTCTCATTGCCCAGCCCAGCTCACGGTGGTGGCACCG
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
410 420 430 440 450 460 470 480 490 500
KC955130_BG8 AGCCTCCGTGTCACTGCCAATGTGGGACAGGACGTTGTGCTGCGCTGCCACTTGTCCCCATGCAAGGATGTTCGGAATTCAGACATCAGATGGATCCAGC
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
510 520 530 540 550 560 570 580 590 600
KC955130_BG8 AGCGGTCCTCTCGGCTTGTGCACCACTACCGAAATGGAGTGGACCTGGGGCAGATGGAGGAATATAAAGGGAGAACAGAACTGCTCAGGGATGGTCTCTC
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)

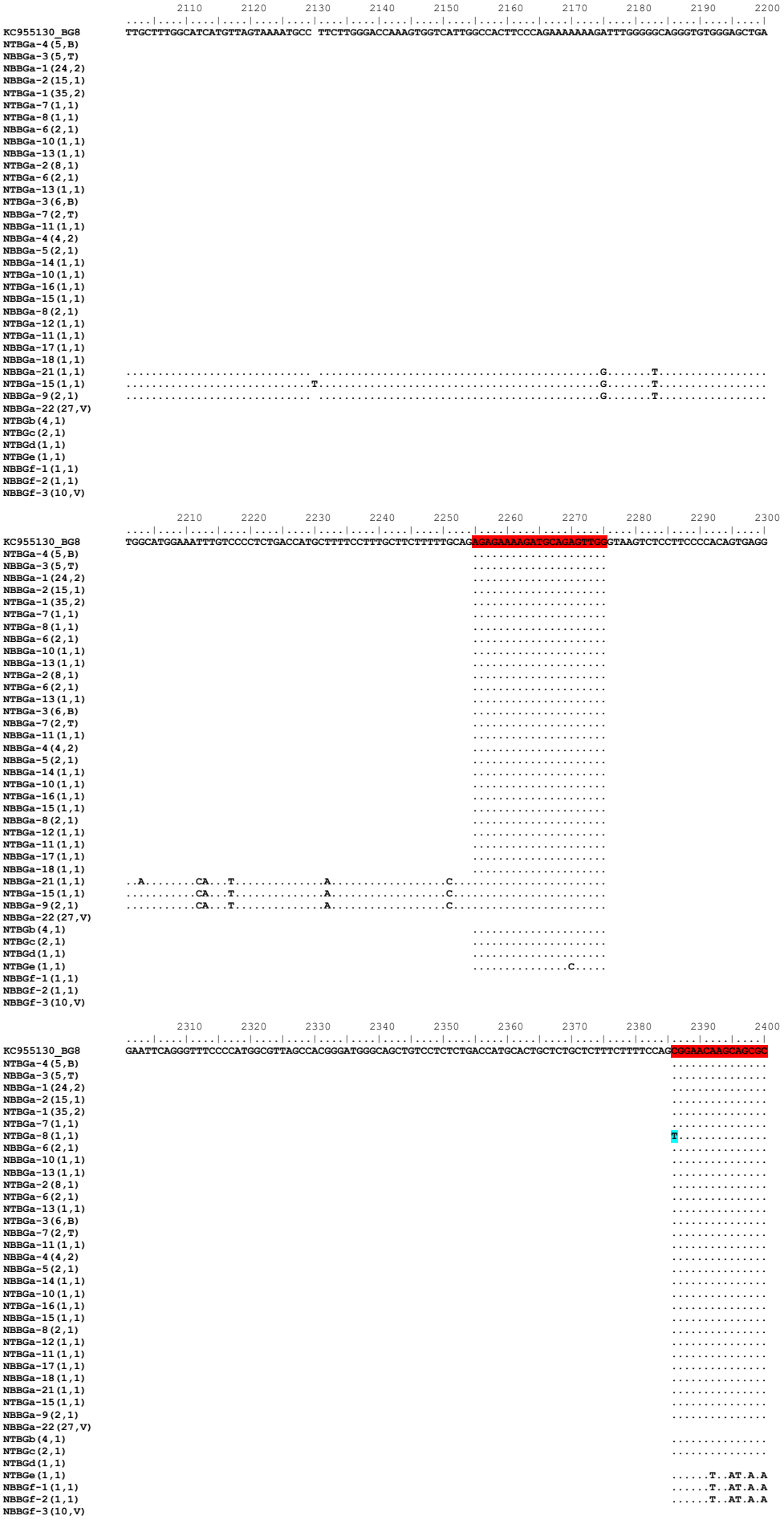

610 620 630 640 650 660 670 680 690 700
KC955130_BG8 TGATGGAAACCTGGATTTGCGCATCACTGCCGTGACCTCCTCTGATAGTGGCTCCTACAGCTGTGCTGTGCAAGATGGTGATGCCTATGCAGAAGCTGTG
NTBGa-4(5,B) ...T...
NBBGa-3(5,T) ...T...
NBBGa-1(24,2) ...T...
NBBGa-2(15,1) ...T...
NTBGa-1(35,2) ...T...
NTBGa-7(1,1) ...T...
NTBGa-8(1,1) ...T...
NBBGa-6(2,1) ...T...
NBBGa-10(1,1) ...T...
NBBGa-13(1,1) ...T...
NTBGa-2(8,1) ...T...
NTBGa-6(2,1) ...T...
NTBGa-13(1,1) ...T...
NTBGa-3(6,B) ...T...
NBBGa-7(2,T) ...T...
NBBGa-11(1,1) ...T...
NBBGa-4(4,2) ...T...
NBBGa-5(2,1) ...T...
NBBGa-14(1,1) ...T...
NTBGa-10(1,1) ...T...
NTBGa-16(1,1) ...T...
NBBGa-15(1,1) ...T...
NBBGa-8(2,1) ...T...
NTBGa-12(1,1) ...T...
NTBGa-11(1,1) ...T...
NBBGa-17(1,1) ...T...
NBBGa-18(1,1) ...T...
NBBGa-21(1,1) ...T...
NTBGa-15(1,1) ...T...
NBBGa-9(2,1) ...T...
NBBGa-22(27,V) ...T...
NTBGb(4,1) ...T...
NTBGc(2,1) ...T.....C...
NTBGd(1,1) ...T...
NTBGe(1,1) ...T...G.A...C.....A.....G.....G.....C.....
NBBGf-1(1,1) ...T...G.A...C.....A.....G.....G.....C.....
NBBGf-2(1,1) ...T...G.A...C.....A.....G.....G.....C.....
NBBGf-3(10,V) ...T...G.A...C.....A.....G.....G.....C.....
710 720 730 740 750 760 770 780 790 800
KC955130_BG8 GTGAACCTGGAGGTGTCAGGTCAGTGGCTGGGGTGTTCAAGGATGGAGAGCTGACGGATCGCAGCCTTTGGAAGTGGTCAGGGCTGAACAGCTCCATGAG
NTBGa-4(5,B) .....
NBBGa-3(5,T) .....
NBBGa-1(24,2) .....
NBBGa-2(15,1) .....
NTBGa-1(35,2) .....
NTBGa-7(1,1) .....
NTBGa-8(1,1) .....
NBBGa-6(2,1) .....
NBBGa-10(1,1) .....
NBBGa-13(1,1) .....
NTBGa-2(8,1) .....
NTBGa-6(2,1) .....
NTBGa-13(1,1) .....
NTBGa-3(6,B) .....
NBBGa-7(2,T) .....
NBBGa-11(1,1) .....
NBBGa-4(4,2) .....
NBBGa-5(2,1) .....
NBBGa-14(1,1) .....
NTBGa-10(1,1) .....
NTBGa-16(1,1) .....
NBBGa-15(1,1) .....
NBBGa-8(2,1) .....
NTBGa-12(1,1) .....
NTBGa-11(1,1) .....
NBBGa-17(1,1) .....
NBBGa-18(1,1) .....
NBBGa-21(1,1) .....
NTBGa-15(1,1) .....
NBBGa-9(2,1) .....
NBBGa-22(27,V) .....
NTBGb(4,1) .....
NTBGc(2,1) .....
NTBGd(1,1) .....
NTBGe(1,1) ...G.....
NBBGf-1(1,1) ...G.....
NBBGf-2(1,1) ...G.....
NBBGf-3(10,V) ...G.....
810 820 830 840 850 860 870 880 890 900
KC955130_BG8 ATGCTGGAATTGCAGTGGGCGCACGCTGTGATTTGGAGATGGGTCTGCATGGATGAGGTGGTTGGGTTGGGTTTCTGGGATGGGTTTCTCCATGGCTCAG
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)

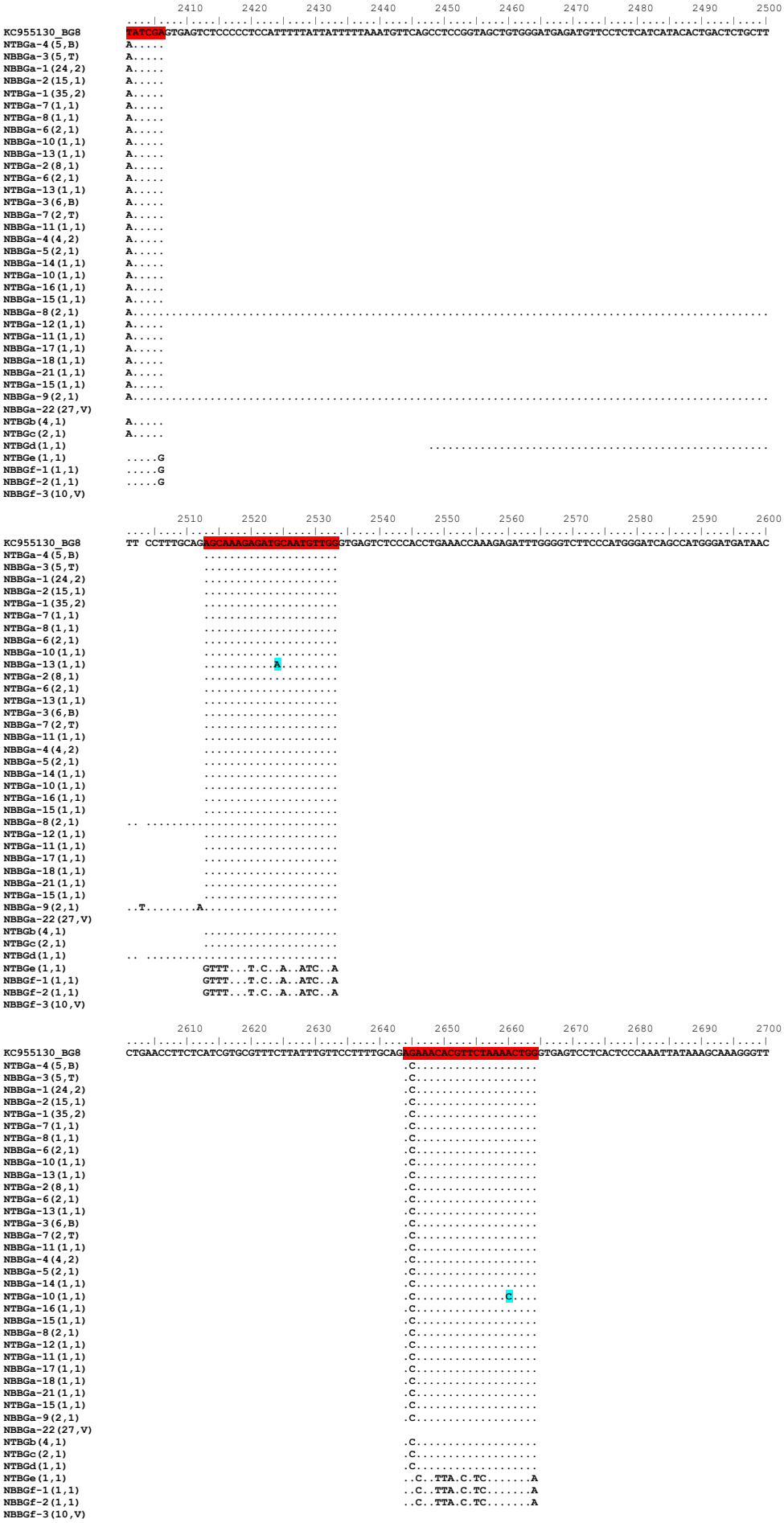

910 920 930 940 950 960 970 980 990 1000
KC955130_BG8 TGGCAGTCGGCACACAATGCTGAGCAGCTCCTCTGCCTGTGCCAATGTGGGGATGCTGCTATTGTGTGTCACTGCTCGCTGGTTGCTGCCCCTTCGGGTT
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
KC955130_BG8 CTGTGATCTCCCAAGGCTGAGTCTTGCTTTTCCACATATGGGAATTTAAAAGGACCCTCTTCTTGACATTTCTTCCAGACCCCTTTTCTATGATCATCCT
NTBGa-4(5,B) .....
NBBGa-3(5,T) .....
NBBGa-1(24,2) .....
NBBGa-2(15,1) .....
NTBGa-1(35,2) .....
NTBGa-7(1,1) .....
NTBGa-8(1,1) .....
NBBGa-6(2,1) .....
NBBGa-10(1,1) .....
NBBGa-13(1,1) .....
NTBGa-2(8,1) .....
NTBGa-6(2,1) .....
NTBGa-13(1,1) .....
NTBGa-3(6,B) .....
NBBGa-7(2,T) .....
NBBGa-11(1,1) .....
NBBGa-4(4,2) .....
NBBGa-5(2,1) .....
NBBGa-14(1,1) .....
NTBGa-10(1,1) .....
NTBGa-16(1,1) .....
NBBGa-15(1,1) .....
NBBGa-8(2,1) .....
NTBGa-12(1,1) .....
NTBGa-11(1,1) .....
NBBGa-17(1,1) .....
NBBGa-18(1,1) .....
NBBGa-21(1,1) .....
NTBGa-15(1,1) .....
NBBGa-9(2,1) .....
NBBGa-22(27,V)
NTBGb(4,1) .....
NTBGc(2,1) .....
NTBGd(1,1) .....
NTBGe(1,1) .T.....CCA....G...A
NBBGf-1(1,1) .T.....CCA....G...A
NBBGf-2(1,1) .T.....CCA....G...A
NBBGf-3(10,V)
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
KC955130_BG8 TTACTGGACAGTGGCTCTGGCTGTGATCATCACACTTCTGGTTGGGTCATTTGTCGTCAATGTTTTT CTCCATAGAAAGAAAAGGTGAGCTGAGAGCGGA
NTBGa-4(5,B) ....
NBBGa-3(5,T) ....
NBBGa-1(24,2) ....
NBBGa-2(15,1) ....
NTBGa-1(35,2) ....
NTBGa-7(1,1) ....
NTBGa-8(1,1) ....
NBBGa-6(2,1) .....A..... .....
NBBGa-10(1,1) ....
NBBGa-13(1,1) ....
NTBGa-2(8,1) .....T..... .....
NTBGa-6(2,1) ....
NTBGa-13(1,1) ....
NTBGa-3(6,B) ....
NBBGa-7(2,T) ....
NBBGa-11(1,1) ....
NBBGa-4(4,2) ....
NBBGa-5(2,1) ....
NBBGa-14(1,1) ....
NTBGa-10(1,1) ....
NTBGa-16(1,1) ....
NBBGa-15(1,1) ....
NBBGa-8(2,1) ....
NTBGa-12(1,1) .....T..... .....
NTBGa-11(1,1) ....
NBBGa-17(1,1) ....
NBBGa-18(1,1) ....
NBBGa-21(1,1) ....
NTBGa-15(1,1) ....
NBBGa-9(2,1) ....
NBBGa-22(27,V)
NTBGb(4,1) ....
NTBGc(2,1) ....
NTBGd(1,1) ....
NTBGe(1,1) .CC.....AG.....G..G.....A....C.....A..... ...TG...G.....
NBBGf-1(1,1) .CC.....AG.....G..G.....A....C.....A..... ...TG...G.....
NBBGf-2(1,1) .CC.....AG.....G..G.....A....C.....A..... ...TG...G.....
NBBGf-3(10,V)

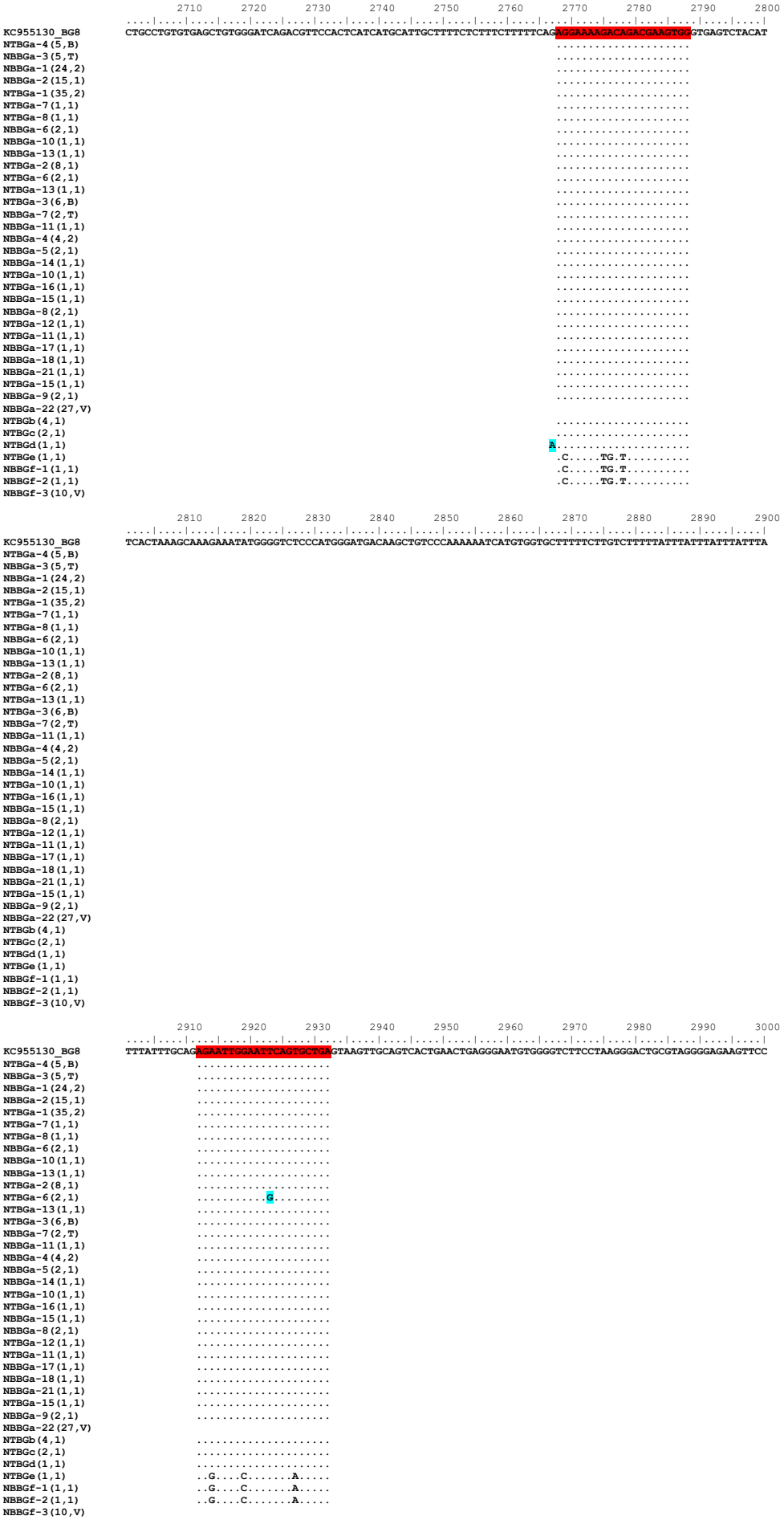
```
                                1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KC955130_BG8       GGGGATGGAGCACAGGGAGGTGTTGTGCATGGACAGGGATGGTCGGGGTGGTGCTGAGCTGTGGTCCACGGAGGTACACAGGTGGAGGAACCGTGACTTT
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)

                                1310      1320      1330      1340      1350      1360      1370      1380      1390      1400
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KC955130_BG8       TCATGGGATTCCCAGTGCTCATTAAATAACATTTGCCTTTCTTTTGGGGAATAAAAGAAGGGGAAAAACGATAGTGGTAAGGGTGGGCAGATAGGAATGT
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)

                                1410      1420      1430      1440      1450      1460      1470      1480      1490      1500
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
KC955130_BG8       GGCTGGACTGTGGGGCAGGTGGAAAGTCCAAACCCTCTGGAGAAGTCCCCACAAACCAAGCTGCCCTGCTGACCAGCTATTTCTCTGCTTTGTTTTCCAG
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
```

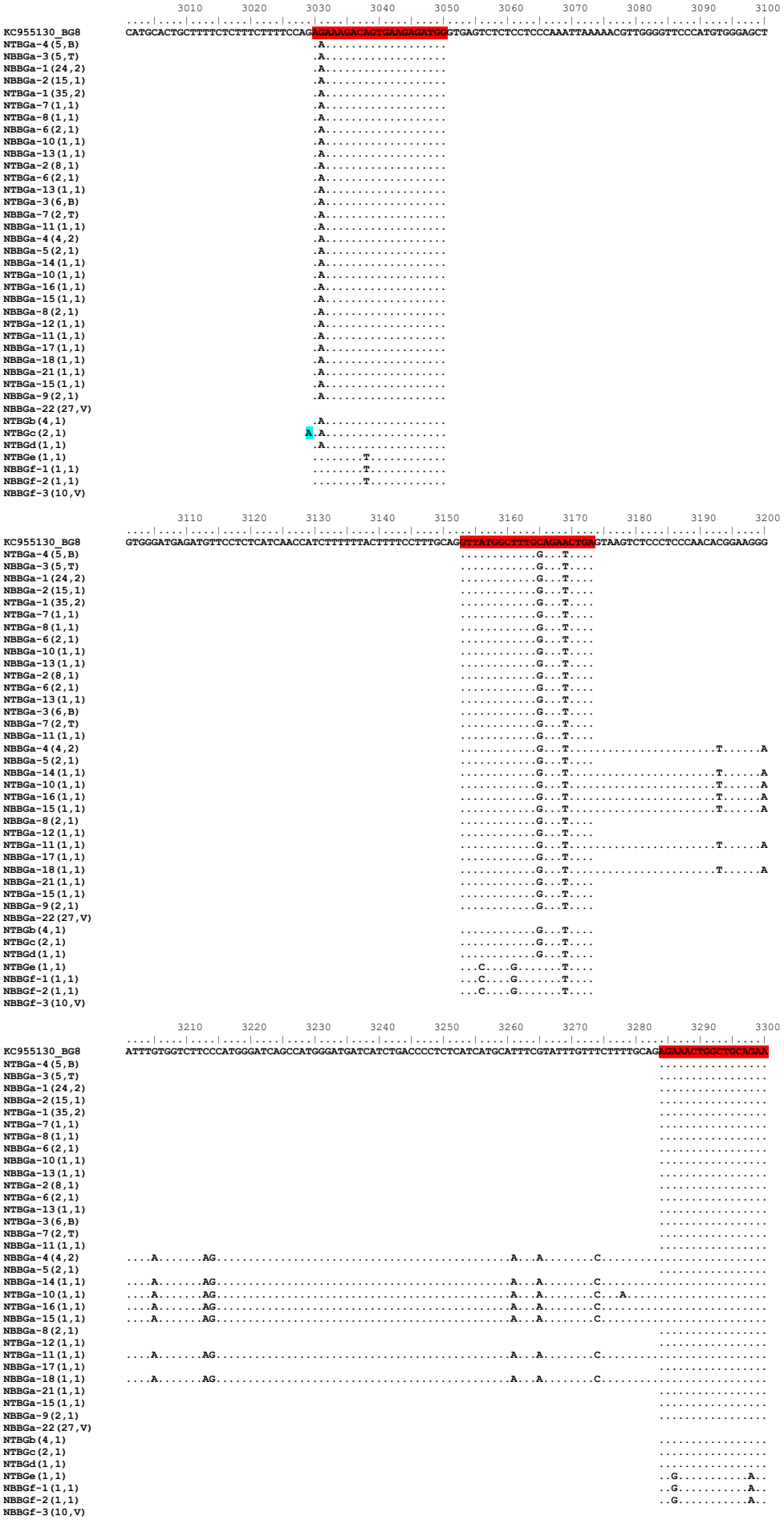

1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
KC955130_BG8 TGGCACAGAGCAGAGAGCTGAGTGAGTCCTTCCATCCCCATCCACCACCAAAGTCCCTTTAATGGAACTGACAGCAGACTGCAGAGTGCTGGGTTATGCC
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1) A
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1) C
NBBGf-1(1,1) C G T T..A
NBBGf-2(1,1) C
NBBGf-3(10,V)
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
KC955130_BG8 ATGTGCTGGGGCCATGAGCTATGTTGAGGCTTTGGAATGTGTTGGGGTTGTGGGATGTACTGGGGTCGTGGGATGTGTCAATCCTGGCTGATTCACGTGG
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1) T.T
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
KC955130_BG8 AAAAACCTTTCACAATCGGTTCCTTCCAGTTTGTTTAATTCCTTCTTGGGCCCAAAGTGGTCATTGGACTCCTCCCAGAAAAAAGGGTTTGGGGTCAGGG
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)

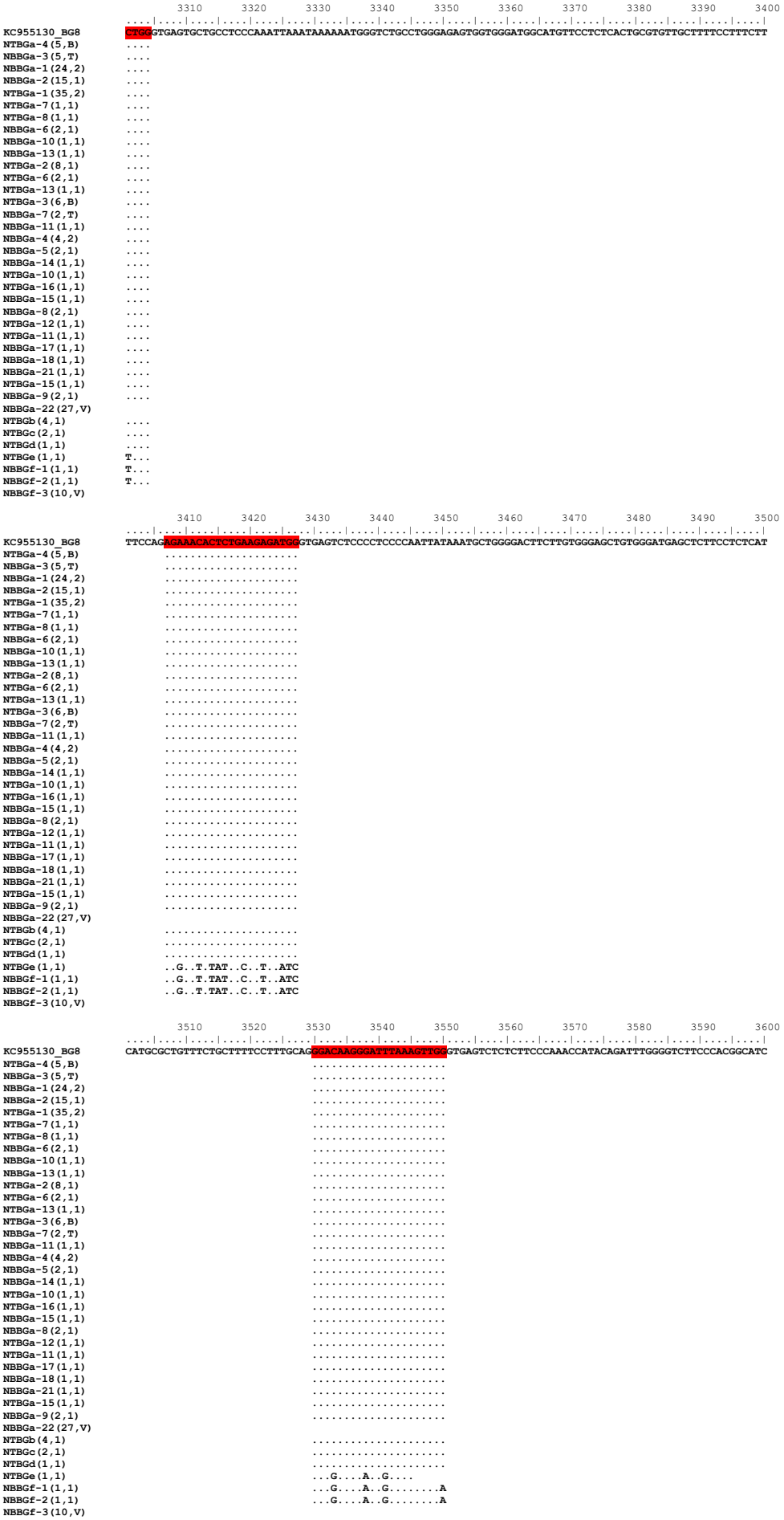

1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
KC955130_BG8 TGTGAGAGCTGATGGCATGGAAACGTGTCCCCTCTGACCATGCATTTCATTTGCTTCTATTTTGCAGAGAGAAAAGATGCAGAGTTGGGTAAGTCTCCTT
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T) ...A.....
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1) .....T.....
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1) .....C.....
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1) .....C.....
NBBGf-2(1,1) .....C.....
NBBGf-3(10,V)
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
KC955130_BG8 CCCTAAAGCGAGGGAATTCAGGGTGTCCCCATGGCATCA GCCGTGGAATTAGTAGCTGTCCTCTCTGACAATTCACTGCTCTGCTCTTTCCTTTCCAGT
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1) .....GC.GTG.A.....
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
KC955130_BG8 GGAGAAAGCTGCAGCATTGGGTGAGTTATATTCCCCAAGCCAAAGTACTTTGGGTCTTCCCATTGGAAGTTATTTCCTCAGACCATCCTTTCTGTTGTGT
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)

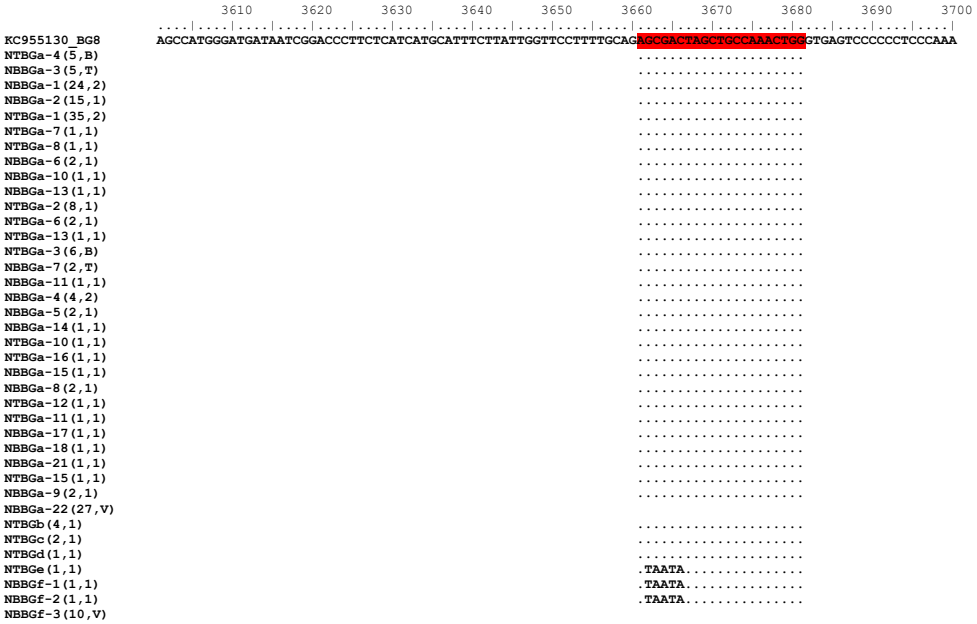

2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
KC955130_BG8 TTGCTTTGGCATCATGTTAGTAAAATGCC TTCTTGGGACCAAAGTGGTCATTGGCCACTTCCCAGAAAAAAAGATTTGGGGGCAGGGTGTGGGAGCTGA
2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
KC955130_BG8 TGGCATGGAAATTTGTCCCCTCTGACCATGCTTTTCCTTTGCTTCTTTTTGCAGAGAGAAAAGATGCAGAGTTGGGTAAGTCTCCTTCCCCACAGTGAGG
2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
KC955130_BG8 GAATTCAGGGTTTCCCCATGGCGTTAGCCACGGGATGGGCAGCTGTCCTCTCTGACCATGCACTGCTCTGCTCTTTCTTTTCCAGCGGAACAAGCAGCGC

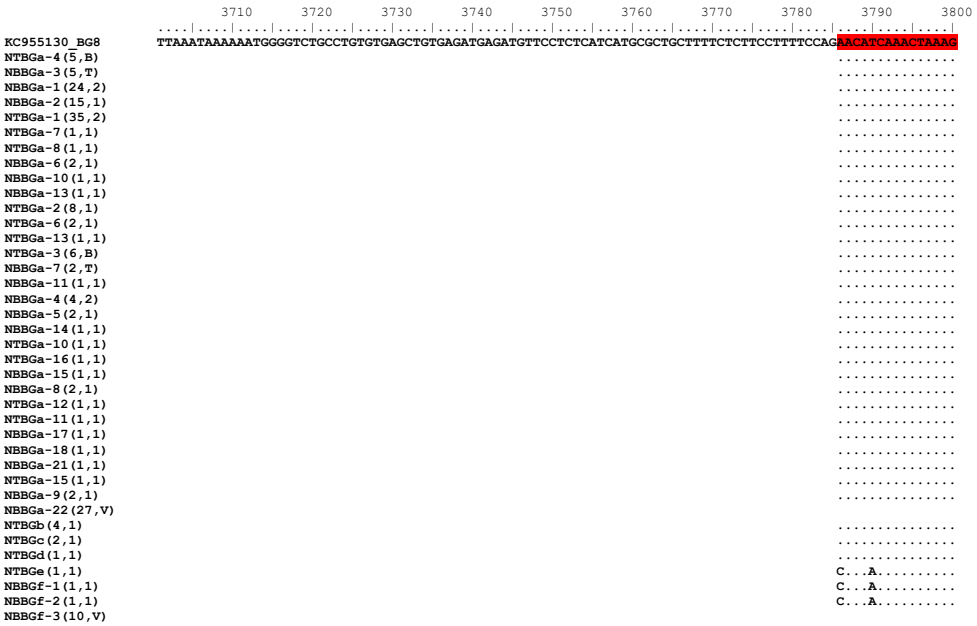

2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
KC955130_BG8 TATCGAGTGAGTCTCCCCCTCCATTTTTATTATTTTTAAATGTTCAGCCTCCGGTAGCTGTGGGATGAGATGTTCCTCTCATCATACACTGACTCTGCTT
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1) .....G
NBBGf-1(1,1) .....G
NBBGf-2(1,1) .....G
NBBGf-3(10,V)
2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
KC955130_BG8 TT CCTTTGCAGAGCAAAGAGATGCAATGTTGGGTGAGTCTCCCACCTGAAACCAAAGAGATTTGGGGTCTTCCCATGGGATCAGCCATGGGATGATAAC
NBBGa-13(1,1) A
2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
KC955130_BG8 CTGAACCTTCTCATCGTGCGTTTCTTATTTGTTCCTTTTGCAGAGAAACACGTTCTAAAACTGGGTGAGTCCTCACTCCCAAATTATAAAGCAAAGGGTT
NTBGa-10(1,1) C
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1) ..C..TTA.C.TC.....A
NBBGf-1(1,1) ..C..TTA.C.TC.....A
NBBGf-2(1,1) ..C..TTA.C.TC.....A
NBBGf-3(10,V)

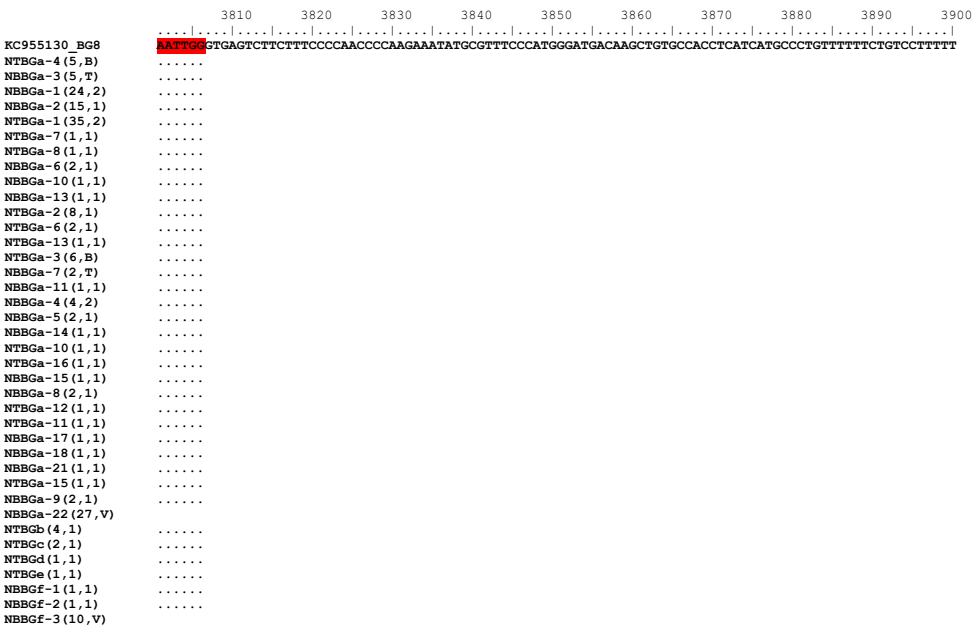

2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
KC955130_BG8 CTGCCTGTGTGAGCTGTGGGATCAGACGTTCCACTCATCATGCATTGCTTTTCTCTTTCTTTTTCAGAGGAAAAGACAGACGAAGTGGGTGAGTCTACAT
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
2810 2820 2830 2840 2850 2860 2870 2880 2890 2900
KC955130_BG8 TCACTAAAGCAAAGAAATATGGGGTCTCCCATGGGATGACAAGCTGTCCCAAAAAATCATGTGGTGCTTTTTCTTGTCTTTTTATTTATTTATTTATTA
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
KC955130_BG8 TTTATTTGCAGAGAATTGGAATTCAGTGCTGAGTAAGTTGCAGTCACTGAACTGAGGGAATGTGGGGTCTTCCTAAGGGACTGCGTAGGGGAGAAGTTCC
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)

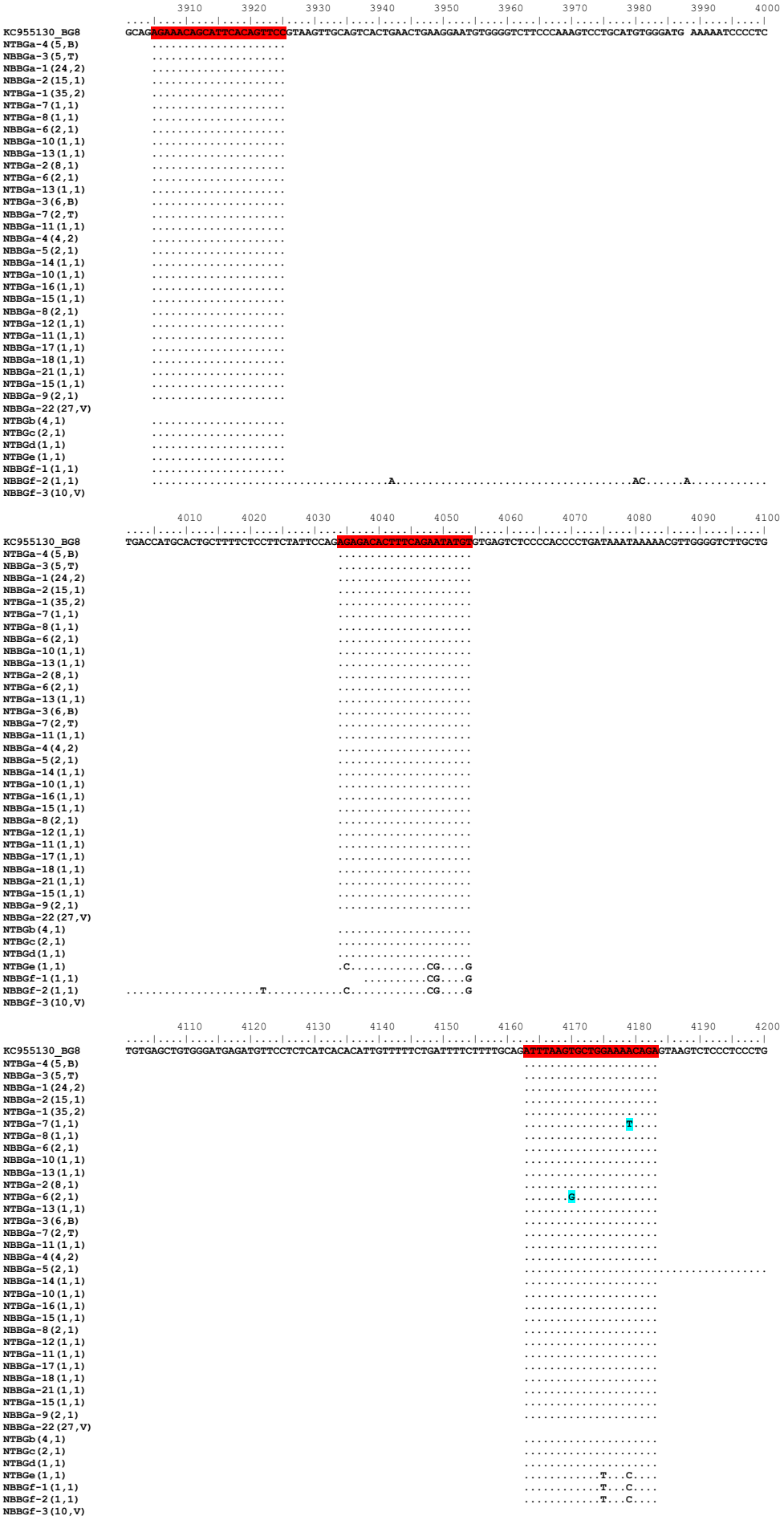

3010 3020 3030 3040 3050 3060 3070 3080 3090 3100
KC955130_BG8 CATGCACTGCTTTTCTCTTTCTTTTCCAGAGAAAGACAGTGAAGAGATGGGTGAGTCTCTCCTCCCAAATTAAAAACGTTGGGGTTCCCATGTGGGAGCT
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
3110 3120 3130 3140 3150 3160 3170 3180 3190 3200
KC955130_BG8 GTGGGATGAGATGTTCCTCTCATCAACCATCTTTTTTACTTTTCCTTTGCAGGTTATGGCTTTGCAGAACTGAGTAAGTCTCCCTCCCAACACGGAAGGG
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
3210 3220 3230 3240 3250 3260 3270 3280 3290 3300
KC955130_BG8 ATTTGTGGTCTTCCCATGGGATCAGCCATGGGATGATCATCTGACCCCTCTCATCATGCATTTCGTATTTGTTTCTTTTGCAGAGAAACTGGCTGCAGAA
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)

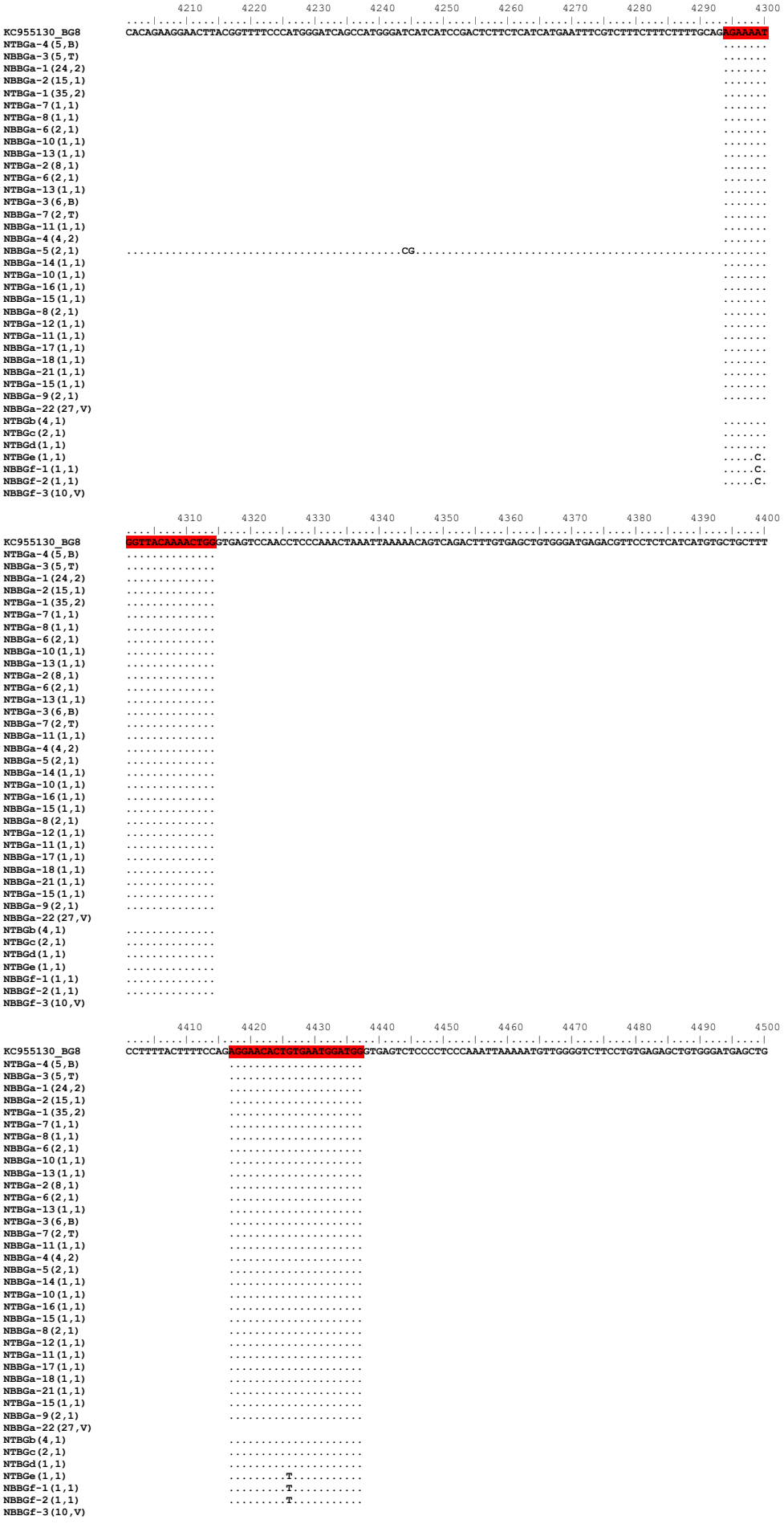

3310 3320 3330 3340 3350 3360 3370 3380 3390 3400
KC955130_BG8 CTGGGTGAGTGCTGCCTCCCAAATTAAATAAAAAATGGGTCTGCCTGGGAGAGTGGTGGGATGGCATGTTCCTCTCACTGCGTGTTGCTTTTCCTTTCTT
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1) T...
NBBGf-1(1,1) T...
NBBGf-2(1,1) T...
NBBGf-3(10,V)
3410 3420 3430 3440 3450 3460 3470 3480 3490 3500
KC955130_BG8 TTCCAGAGAAACACTCTGAAGAGATGGGTGAGTCTCCCCTCCCCAATTATAAATGCTGGGGACTTCTTGTGGGAGCTGTGGGATGAGCTCTTCCTCTCAT
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1) ..G..T.TAT..C..T..ATC
NBBGf-1(1,1) ..G..T.TAT..C..T..ATC
NBBGf-2(1,1) ..G..T.TAT..C..T..ATC
NBBGf-3(10,V)
3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
KC955130_BG8 CATGCGCTGTTTCTGCTTTTCCTTTGCAGGGACAAGGGATTTAAAGTTGGGTGAGTCTCTCTTCCCAAACCATACAGATTTGGGGTCTTCCCACGGCATC
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1) ...G....A..G....
NBBGf-1(1,1) ...G....A..G........A
NBBGf-2(1,1) ...G....A..G........A
NBBGf-3(10,V)

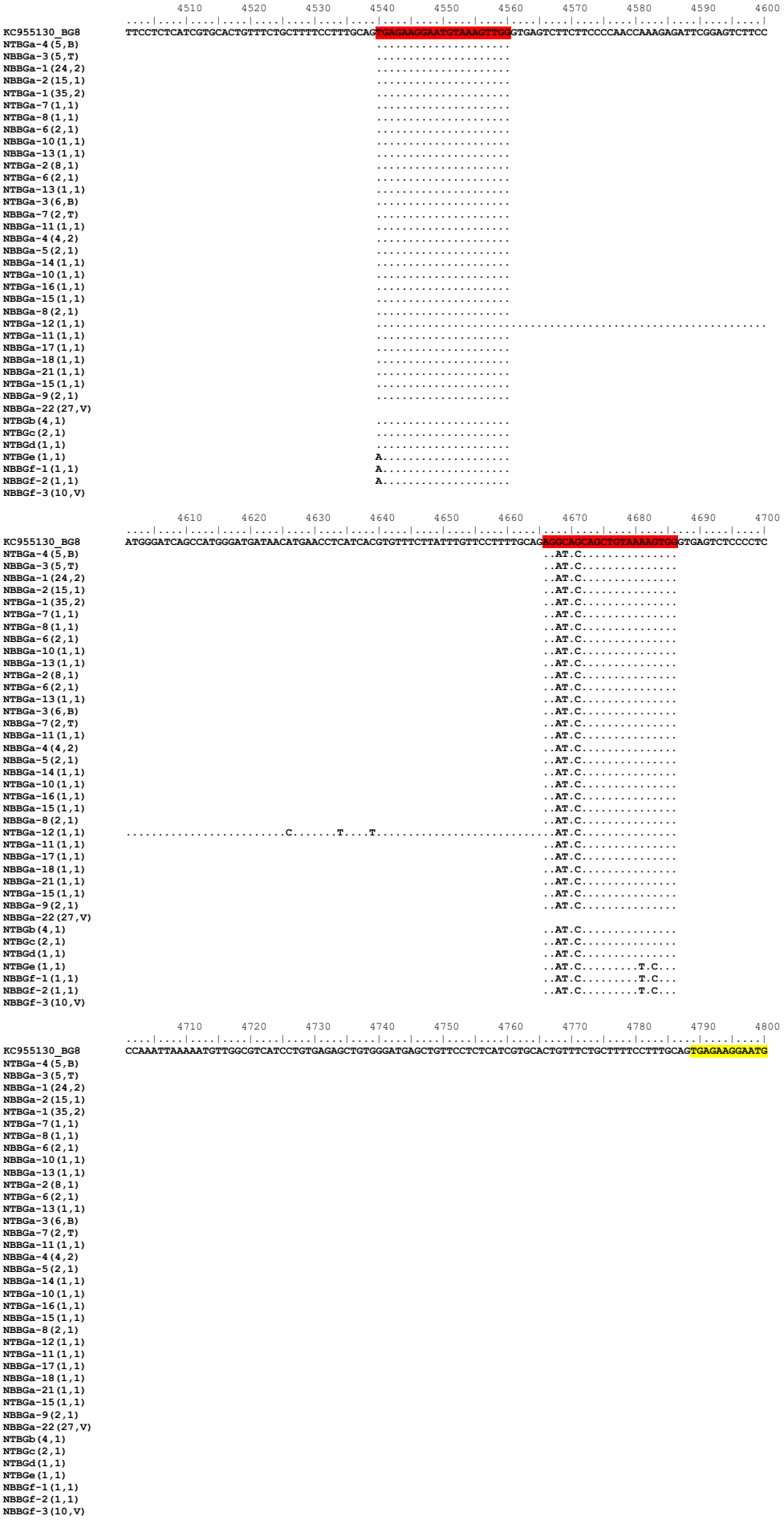

3610 3620 3630 3640 3650 3660 3670 3680 3690 3700
KC955130_BG8 AGCCATGGGATGATAATCGGACCCTTCTCATCATGCATTTCTTATTGGTTCCTTTTGCAGAGCGACTAGCTGCCAAACTGGGTGAGTCCCCCCTCCCAAA
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1) .TAATA.....
NBBGf-1(1,1) .TAATA.....
NBBGf-2(1,1) .TAATA.....
NBBGf-3(10,V)

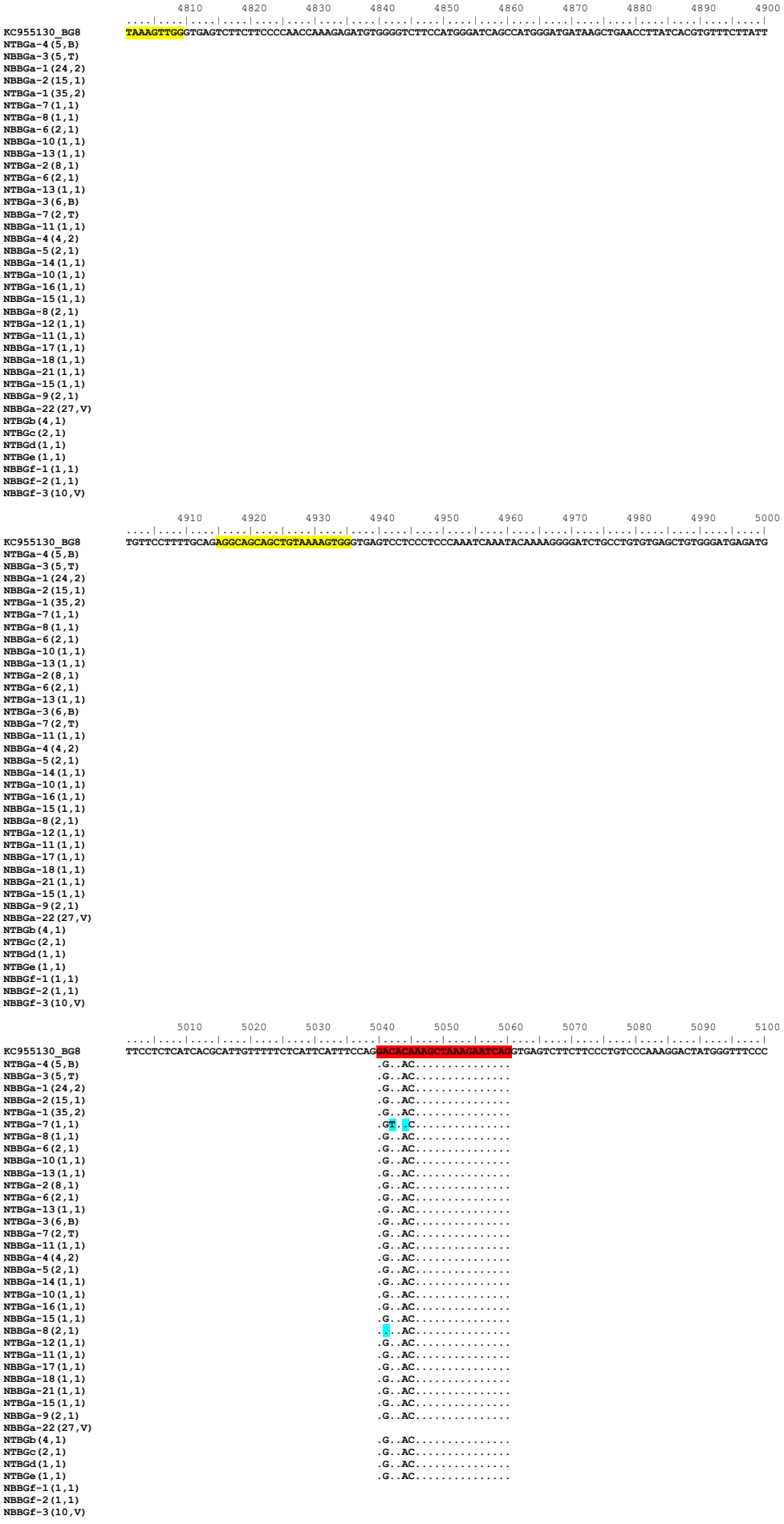

3710 3720 3730 3740 3750 3760 3770 3780 3790 3800
KC955130_BG8 TTAAATAAAAAATGGGGTCTGCCTGTGTGAGCTGTGAGATGAGATGTTCCTCTCATCATGCGCTGCTTTTCTCTTCCTTTTCCAGAACATCAAACTAAAG
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1) C...A.....
NBBGf-1(1,1) C...A.....
NBBGf-2(1,1) C...A.....
NBBGf-3(10,V)

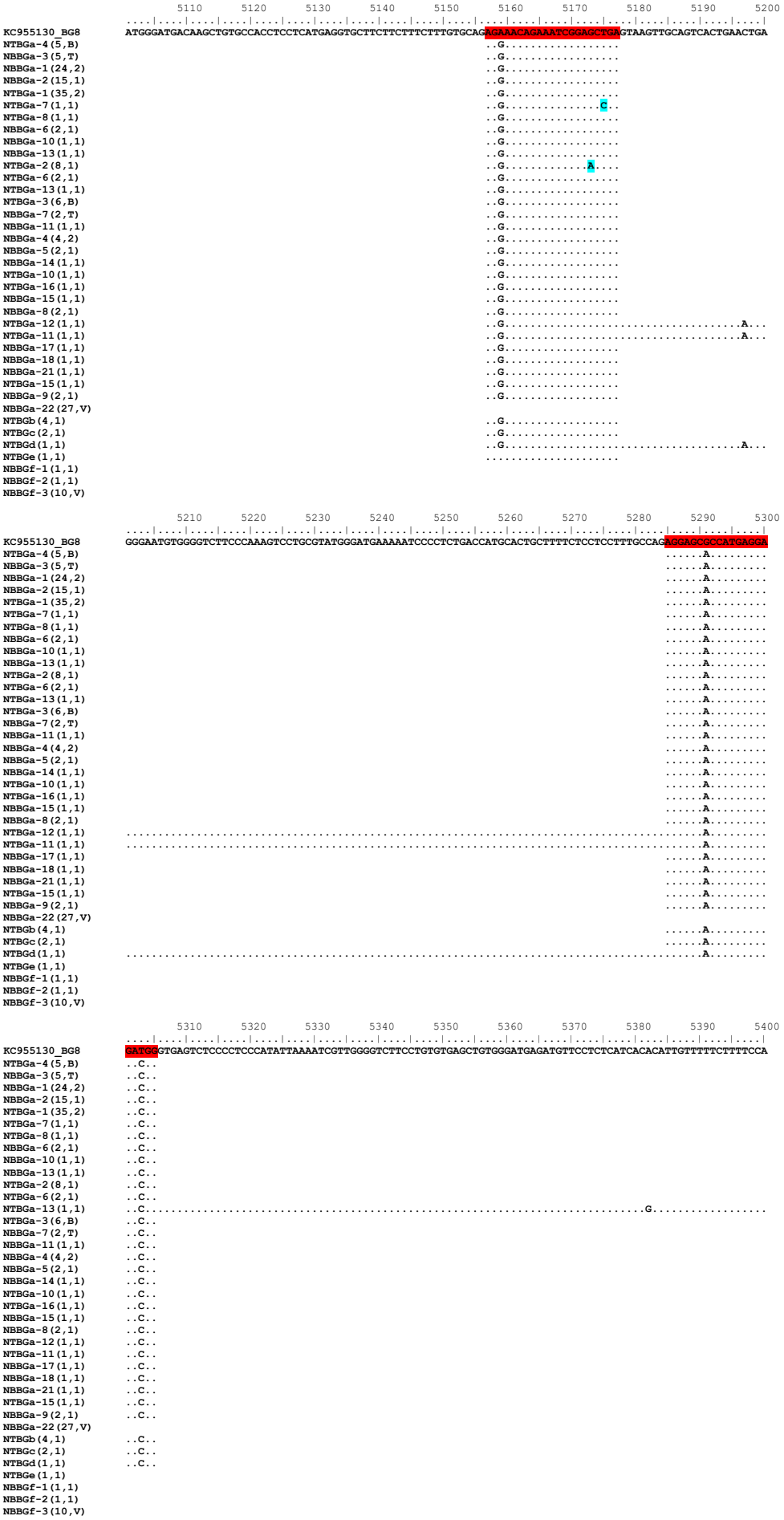

3810 3820 3830 3840 3850 3860 3870 3880 3890 3900
KC955130_BG8 AATTGGGTGAGTCTTCTTTCCCCAACCCCAAGAAATATGCGTTTCCCATGGGATGACAAGCTGTGCCACCTCATCATGCCCTGTTTTTTCTGTCCTTTTT
NTBGa-4(5,B) ......
NBBGa-3(5,T) ......
NBBGa-1(24,2) ......
NBBGa-2(15,1) ......
NTBGa-1(35,2) ......
NTBGa-7(1,1) ......
NTBGa-8(1,1) ......
NBBGa-6(2,1) ......
NBBGa-10(1,1) ......
NBBGa-13(1,1) ......
NTBGa-2(8,1) ......
NTBGa-6(2,1) ......
NTBGa-13(1,1) ......
NTBGa-3(6,B) ......
NBBGa-7(2,T) ......
NBBGa-11(1,1) ......
NBBGa-4(4,2) ......
NBBGa-5(2,1) ......
NBBGa-14(1,1) ......
NTBGa-10(1,1) ......
NTBGa-16(1,1) ......
NBBGa-15(1,1) ......
NBBGa-8(2,1) ......
NTBGa-12(1,1) ......
NTBGa-11(1,1) ......
NBBGa-17(1,1) ......
NBBGa-18(1,1) ......
NBBGa-21(1,1) ......
NTBGa-15(1,1) ......
NBBGa-9(2,1) ......
NBBGa-22(27,V)
NTBGb(4,1) ......
NTBGc(2,1) ......
NTBGd(1,1) ......
NTBGe(1,1) ......
NBBGf-1(1,1) ......
NBBGf-2(1,1) ......
NBBGf-3(10,V)

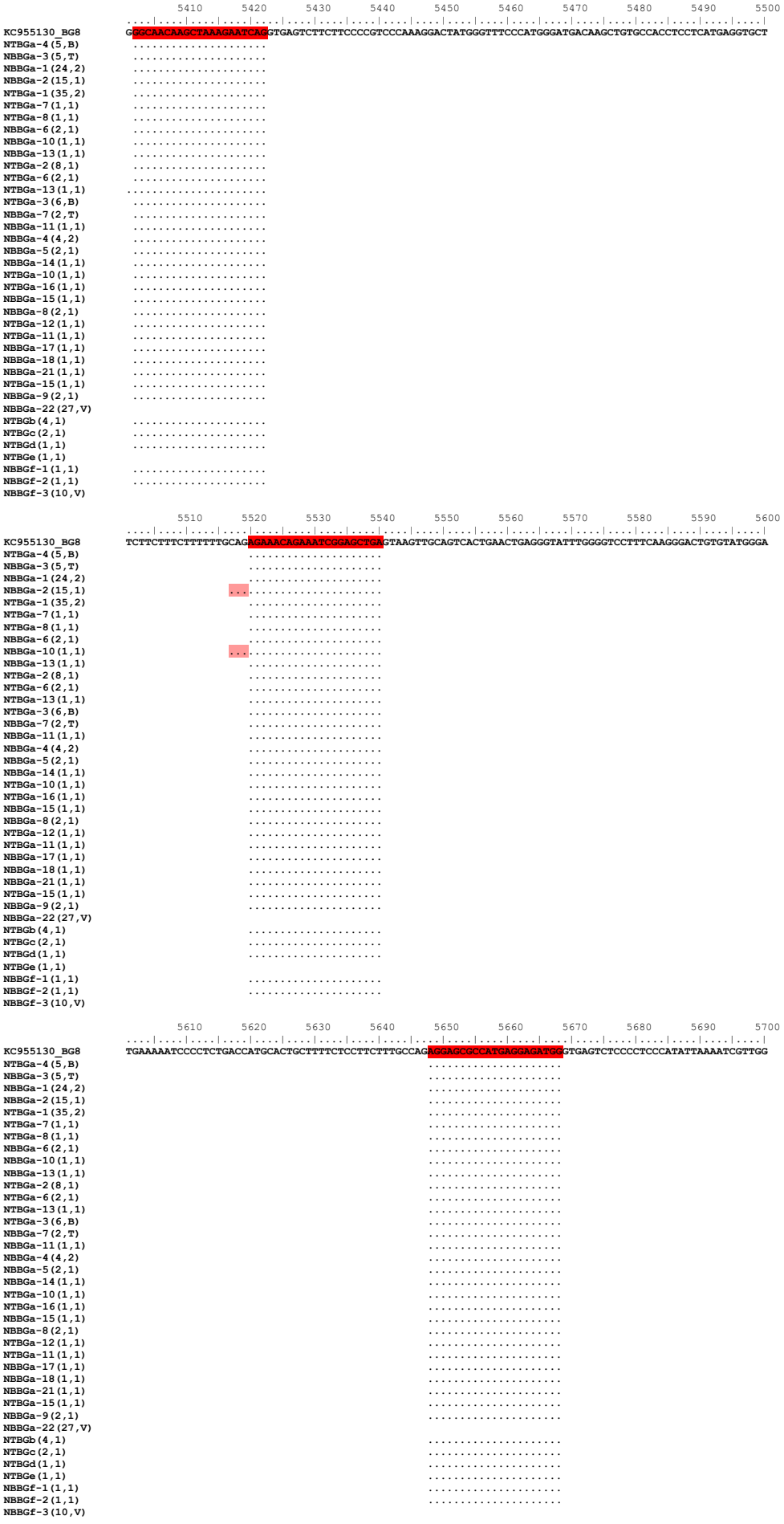

3910 3920 3930 3940 3950 3960 3970 3980 3990 4000
KC955130_BG8 GCAGAGAAACAGCATTCACAGTTCCGTAAGTTGCAGTCACTGAACTGAAGGAATGTGGGGTCTTCCCAAAGTCCTGCATGTGGGATG AAAAATCCCCTC
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1) A AC A
NBBGf-3(10,V)
4010 4020 4030 4040 4050 4060 4070 4080 4090 4100
KC955130_BG8 TGACCATGCACTGCTTTTCTCCTTCTATTCCAGAGAGACACTTTCAGAATATGTGTGAGTCTCCCCACCCCTGATAAATAAAAACGTTGGGGTCTTGCTG
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1) .C.....CG....G
NBBGf-1(1,1) .....CG....G
NBBGf-2(1,1) .....T.....C.....CG....G
NBBGf-3(10,V)
4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
KC955130_BG8 TGTGAGCTGTGGGATGAGATGTTCCTCTCATCACACATTGTTTTTCTGATTTTCTTTTGCAGATTTAAGTGCTGGAAAACAGAGTAAGTCTCCCTCCCTG
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1) T
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1) G
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1) T...C....
NBBGf-1(1,1) T...C....
NBBGf-2(1,1) T...C....
NBBGf-3(10,V)

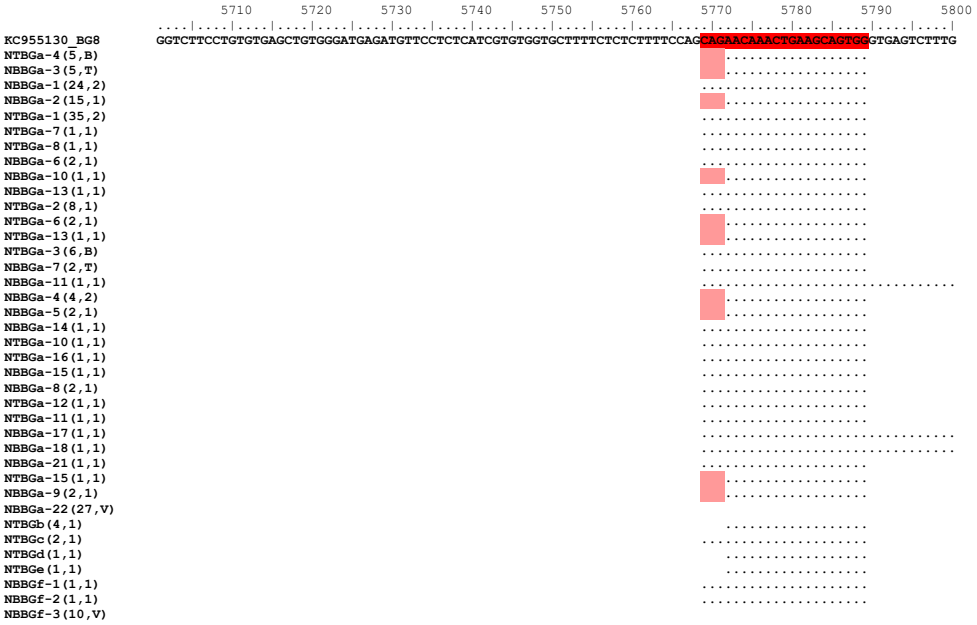

4210 4220 4230 4240 4250 4260 4270 4280 4290 4300
KC955130_BG8 CACAGAAGGAACTTACGGTTTTCCCATGGGATCAGCCATGGGATCATCATCCGACTCTTCTCATCATGAATTTCGTCTTTCTTTCTTTTGCAGAGAAAAT
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1) .....CG.....
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1) .....C.
NBBGf-1(1,1) .....C.
NBBGf-2(1,1) .....C.
NBBGf-3(10,V)
4310 4320 4330 4340 4350 4360 4370 4380 4390 4400
KC955130_BG8 GGTTACAAAACTGGGTGAGTCCAACCTCCCAAACTAAATTAAAAACAGTCAGACTTTGTGAGCTGTGGGATGAGACGTTCCTCTCATCATGTGCTGCTTT
NTBGa-4(5,B) ....
NBBGa-3(5,T) ....
NBBGa-1(24,2) ....
NBBGa-2(15,1) ....
NTBGa-1(35,2) ....
NTBGa-7(1,1) ....
NTBGa-8(1,1) ....
NBBGa-6(2,1) ....
NBBGa-10(1,1) ....
NBBGa-13(1,1) ....
NTBGa-2(8,1) ....
NTBGa-6(2,1) ....
NTBGa-13(1,1) ....
NTBGa-3(6,B) ....
NBBGa-7(2,T) ....
NBBGa-11(1,1) ....
NBBGa-4(4,2) ....
NBBGa-5(2,1) ....
NBBGa-14(1,1) ....
NTBGa-10(1,1) ....
NTBGa-16(1,1) ....
NBBGa-15(1,1) ....
NBBGa-8(2,1) ....
NTBGa-12(1,1) ....
NTBGa-11(1,1) ....
NBBGa-17(1,1) ....
NBBGa-18(1,1) ....
NBBGa-21(1,1) ....
NTBGa-15(1,1) ....
NBBGa-9(2,1) ....
NBBGa-22(27,V)
NTBGb(4,1) ....
NTBGc(2,1) ....
NTBGd(1,1) ....
NTBGe(1,1) ....
NBBGf-1(1,1) ....
NBBGf-2(1,1) ....
NBBGf-3(10,V)
4410 4420 4430 4440 4450 4460 4470 4480 4490 4500
KC955130_BG8 CCTTTTACTTTTCCAGAGGAACACTGTGAATGGATGGGTGAGTCTCCCCTCCCAAATTAAAAATGTTGGGGTCTTCCTGTGAGAGCTGTGGGATGAGCTG
NTBGa-4(5,B) ....
NBBGa-3(5,T) ....
NBBGa-1(24,2) ....
NBBGa-2(15,1) ....
NTBGa-1(35,2) ....
NTBGa-7(1,1) ....
NTBGa-8(1,1) ....
NBBGa-6(2,1) ....
NBBGa-10(1,1) ....
NBBGa-13(1,1) ....
NTBGa-2(8,1) ....
NTBGa-6(2,1) ....
NTBGa-13(1,1) ....
NTBGa-3(6,B) ....
NBBGa-7(2,T) ....
NBBGa-11(1,1) ....
NBBGa-4(4,2) ....
NBBGa-5(2,1) ....
NBBGa-14(1,1) ....
NTBGa-10(1,1) ....
NTBGa-16(1,1) ....
NBBGa-15(1,1) ....
NBBGa-8(2,1) ....
NTBGa-12(1,1) ....
NTBGa-11(1,1) ....
NBBGa-17(1,1) ....
NBBGa-18(1,1) ....
NBBGa-21(1,1) ....
NTBGa-15(1,1) ....
NBBGa-9(2,1) ....
NBBGa-22(27,V)
NTBGb(4,1) ....
NTBGc(2,1) ....
NTBGd(1,1) ....
NTBGe(1,1) .....T.....
NBBGf-1(1,1) .....T.....
NBBGf-2(1,1) .....T.....
NBBGf-3(10,V)

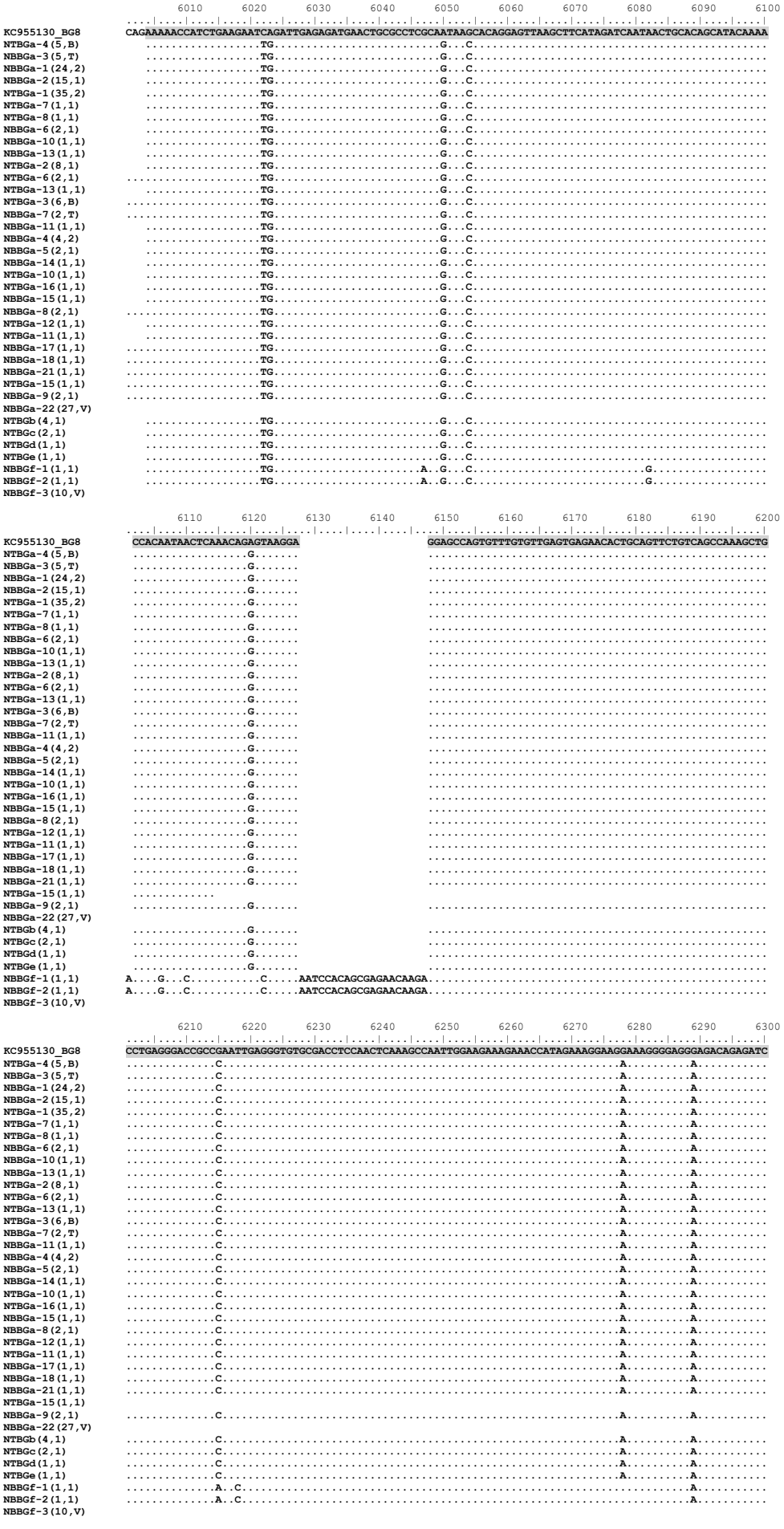

4510 4520 4530 4540 4550 4560 4570 4580 4590 4600
KC955130_BG8 TTCCTCTCATCGTGCACTGTTTCTGCTTTTCCTTTGCAGTGAGAAGGAATGTAAAGTTGGGTGAGTCTTCTTCCCCAACCAAAGAGATTCGGAGTCTTCC
4610 4620 4630 4640 4650 4660 4670 4680 4690 4700
KC955130_BG8 ATGGGATCAGCCATGGGATGATAACATGAACCTCATCACGTGTTTCTTATTTGTTCCTTTTGCAGAGGCAGCAGCTGTAAAAGTGGGTGAGTCTCCCCTC
4710 4720 4730 4740 4750 4760 4770 4780 4790 4800
KC955130_BG8 CCAAATTAAAAATGTTGGCGTCATCCTGTGAGAGCTGTGGGATGAGCTGTTCCTCTCATCGTGCACTGTTTCTGCTTTTCCTTTGCAGTGAGAAGGAATG
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)

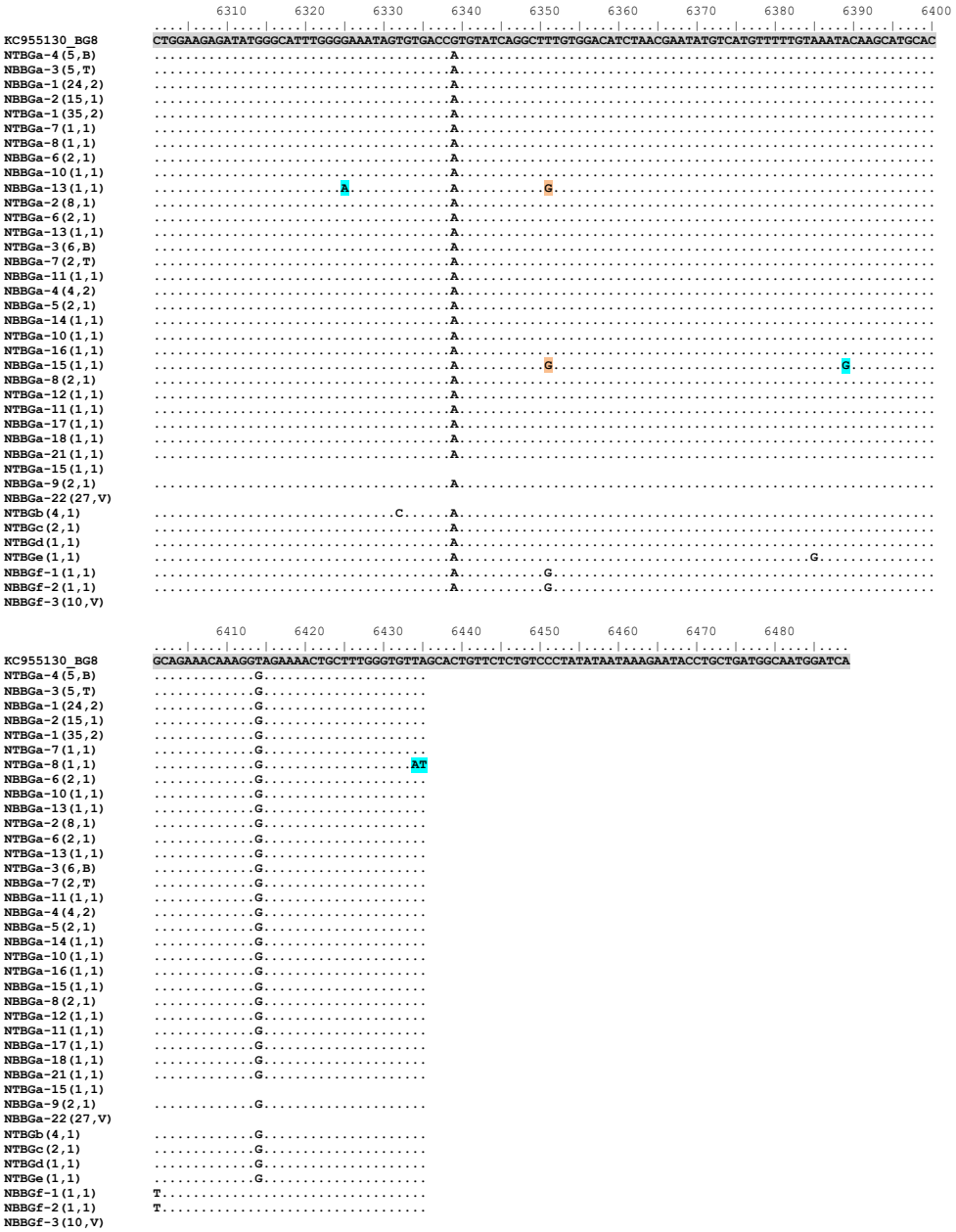

4810 4820 4830 4840 4850 4860 4870 4880 4890 4900
KC955130_BG8 TAAAGTTGGGTGAGTCTTCTTCCCCAACCAAAGAGATGTGGGGTCTTCCATGGGATCAGCCATGGGATGATAAGCTGAACCTTATCACGTGTTTCTTATT
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
4910 4920 4930 4940 4950 4960 4970 4980 4990 5000
KC955130_BG8 TGTTCCTTTTGCAGAGGCAGCAGCTGTAAAAGTGGGTGAGTCCTCCCTCCCAAATCAAATACAAAAGGGGATCTGCCTGTGTGAGCTGTGGGATGAGATG
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
5010 5020 5030 5040 5050 5060 5070 5080 5090 5100
KC955130_BG8 TTCCTCTCATCACGCATTGTTTTTCTCATTCATTTCCAGCACACAAAGCTAAAGAATCAGGTGAGTCTTCTTCCCTGTCCCAAAGGACTATGGGTTTCCC
NTBGa-4(5,B) .G..AC.....
NBBGa-3(5,T) .G..AC.....
NBBGa-1(24,2) .G..AC.....
NBBGa-2(15,1) .G..AC.....
NTBGa-1(35,2) .G..AC.....
NTBGa-7(1,1) .GT..C.....
NTBGa-8(1,1) .G..AC.....
NBBGa-6(2,1) .G..AC.....
NBBGa-10(1,1) .G..AC.....
NBBGa-13(1,1) .G..AC.....
NTBGa-2(8,1) .G..AC.....
NTBGa-6(2,1) .G..AC.....
NTBGa-13(1,1) .G..AC.....
NTBGa-3(6,B) .G..AC.....
NBBGa-7(2,T) .G..AC.....
NBBGa-11(1,1) .G..AC.....
NBBGa-4(4,2) .G..AC.....
NBBGa-5(2,1) .G..AC.....
NBBGa-14(1,1) .G..AC.....
NTBGa-10(1,1) .G..AC.....
NTBGa-16(1,1) .G..AC.....
NBBGa-15(1,1) .G..AC.....
NBBGa-8(2,1) ..G..AC.....
NTBGa-12(1,1) .G..AC.....
NTBGa-11(1,1) .G..AC.....
NBBGa-17(1,1) .G..AC.....
NBBGa-18(1,1) .G..AC.....
NBBGa-21(1,1) .G..AC.....
NTBGa-15(1,1) .G..AC.....
NBBGa-9(2,1) .G..AC.....
NBBGa-22(27,V)
NTBGb(4,1) .G..AC.....
NTBGc(2,1) .G..AC.....
NTBGd(1,1) .G..AC.....
NTBGe(1,1) .G..AC.....
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)


5110 5120 5130 5140 5150 5160 5170 5180 5190 5200
KC955130_BG8 ATGGGATGACAAGCTGTGCCACCTCCTCATGAGGTGCTTCTTCTTTCTTTGTGCAGAGAAACAGAAATCGGAGCTGAGTAAGTTGCAGTCACTGAACTGA
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
5210 5220 5230 5240 5250 5260 5270 5280 5290 5300
KC955130_BG8 GGGAATGTGGGGTCTTCCCAAAGTCCTGCGTATGGGATGAAAAATCCCCTCTGACCATGCACTGCTTTTCTCCTCCTTTGCCAGAGGAGCGCCATGAGGA
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
5310 5320 5330 5340 5350 5360 5370 5380 5390 5400
KC955130_BG8 GATGCGTGAGTCTCCCCTCCCATATTAAAATCGTTGGGGTCTTCCTGTGTGAGCTGTGGGATGAGATGTTCCTCTCATCACACATTGTTTTTCTTTTCCA
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)


5410 5420 5430 5440 5450 5460 5470 5480 5490 5500
KC955130_BG8 GGGCAACAAGCTAAAGAATCAGGTGAGTCTTCTTCCCCGTCCCAAAGGACTATGGGTTTCCCATGGGATGACAAGCTGTGCCACCTCCTCATGAGGTGCT
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
5510 5520 5530 5540 5550 5560 5570 5580 5590 5600
KC955130_BG8 TCTTCTTTCTTTTTTGCAGAGAAACAGAAATCGGAGCTGAGTAAGTTGCAGTCACTGAACTGAGGGTATTTGGGGTCCTTTCAAGGGACTGTGTATGGGA
5610 5620 5630 5640 5650 5660 5670 5680 5690 5700
KC955130_BG8 TGAAAAATCCCCTCTGACCATGCACTGCTTTTCTCCTTCTTTGCCAGAGGAGCGCCATGAGGAGATGGGTGAGTCTCCCCTCCCATATTAAAATCGTTGG


5710 5720 5730 5740 5750 5760 5770 5780 5790 5800
KC955130_BG8 GGTCTTCCTGTGTGAGCTGTGGGATGAGATGTTCCTCTCATCGTGTGGTGCTTTTCTCTCTTTTCCAGCAGAACAAACTGAAGCAGTGGGTGAGTCTTTG
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
5810 5820 5830 5840 5850 5860 5870 5880 5890 5900
KC955130_BG8 TCCCCAAACCAAAGGAATATGGGGCAATCCATGGGATGACAAGCTGTCCCATCTCATCATGTGTTGCTTTTCTATTCCTTTTTT CCAGTGGTAGAAACT
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1) ......G.....G....A.....G.....T.....
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1) ......G.....G....A.....G.....
NBBGa-18(1,1) ......G.....G....A.....G.....
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1) ....T...
NBBGf-2(1,1) ....T...
NBBGf-3(10,V)
5910 5920 5930 5940 5950 5960 5970 5980 5990 6000
KC955130_BG8 GAAGAATAGGGTGAGTCTTTCCCAAACCAAAGCAATACAGGGTTTCCCATGGGATGACAAGCTGTCCCACCTCAGCATCCGTTCCTTTTTATTTCTTTTC
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1) .....G.....C.....C.....
NTBGa-13(1,1)
NTBGa-3(6,B) .....G.....C.....C.....
NBBGa-7(2,T) .....G.....C.....C.....
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1) .....G.....C.....C.....
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1) .....G.....C.....C.....
NBBGa-18(1,1) .....G.....C.....C.....
NBBGa-21(1,1) .....G.....C.....C.....
NTBGa-15(1,1) .....G.....C.....C.....
NBBGa-9(2,1) .....G.....C.....C.....
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1) ......GC..
NBBGf-2(1,1) ......GC..
NBBGf-3(10,V)


6010 6020 6030 6040 6050 6060 6070 6080 6090 6100
KC955130_BG8 CAGAAAAACCATCTGAAGAATCAGATTGAGAGATGAACTGCGCCTCGCAATAAGCACAGGAGTTAAGCTTCATAGATCAATAACTGCACAGCATACAAAA
NTBGa-4(5,B)
NBBGa-3(5,T)
NBBGa-1(24,2)
NBBGa-2(15,1)
NTBGa-1(35,2)
NTBGa-7(1,1)
NTBGa-8(1,1)
NBBGa-6(2,1)
NBBGa-10(1,1)
NBBGa-13(1,1)
NTBGa-2(8,1)
NTBGa-6(2,1)
NTBGa-13(1,1)
NTBGa-3(6,B)
NBBGa-7(2,T)
NBBGa-11(1,1)
NBBGa-4(4,2)
NBBGa-5(2,1)
NBBGa-14(1,1)
NTBGa-10(1,1)
NTBGa-16(1,1)
NBBGa-15(1,1)
NBBGa-8(2,1)
NTBGa-12(1,1)
NTBGa-11(1,1)
NBBGa-17(1,1)
NBBGa-18(1,1)
NBBGa-21(1,1)
NTBGa-15(1,1)
NBBGa-9(2,1)
NBBGa-22(27,V)
NTBGb(4,1)
NTBGc(2,1)
NTBGd(1,1)
NTBGe(1,1)
NBBGf-1(1,1)
NBBGf-2(1,1)
NBBGf-3(10,V)
6110 6120 6130 6140 6150 6160 6170 6180 6190 6200
KC955130_BG8 CCACAATAACTCAAACAGAGTAAGGA GGAGCCAGTGTTTGTGTTGAGTGAGAACACTGCAGTTCTGTCAGCCAAAGCTG
6210 6220 6230 6240 6250 6260 6270 6280 6290 6300
KC955130_BG8 CCTGAGGGACCGCCGAATTGAGGGTGTGCGACCTCCAACTCAAAGCCAATTGGAAGAAAGAAACCATAGAAAGGAAGGAAAGGGGAGGGAGACAGAGATC



Appendix K. Alignment of all the cDNA for five BG genes found in T and B cells from line N (B21) against the BG8 gene (including introns) of B12 haplotype. Name of the gene on the top follows the convention: GenBank accession number of B12 BG region genomic sequence, a dash, "BG" and the number of the gene locus for the B12 haplotype. Names of the transcripts follow the convention: "N" for line N, "T" for T cells, "B" for B cells, "BG" and the letter representing the exon 2 sequence. Numbers in parenthesis indicate the number of times the depicted sequence was found out of the number of times the gene (based on the exon 2 sequence) was found, followed by the number of independent PCRs in which the sequence was identified (1, found in one PCR; 2 found in 2 PCRs; T, found in one PCR but also in at least one PCR in B cell cDNA; B, found in one PCR PCR but also in at least one PCR in T cell cDNA; V, found using SS-TM primers which only containing partial signal sequence, the whole Ig-V domain and partial transmembrane region). Colours indicate different coding regions (grey, 5'UTR and 3'UTR; dark green, signal sequence; light green, Ig-V domain; brown, transmembrane region; alternating yellow and red, cytoplasmic tail regions (codons from each 18, 21 or 24 nucleotide repeat, with nucleotides in the split codon indicated by the color of the exon in which the majority of the codon is located); purple, in-frame stop codon; light blue, introns (including two positions that are probably nucleotide mis-incorporations during PCR reaction) which are all deleted in the analyses for "(nearly) full-length conceptual transcripts" (that is, exons without introns). Letters indicate nucleotides, dot indicates identity with BG8 sequence; dash indicates not present in all sequences.

Appendix L. Summary of all BG cDNA sequences found from four chicken lines

(L1). All BG cDNA sequences found in B cells of line N (B21)

Line N (B21)	Representative Clones	Original Names	Note	Final Names
NBBGa(97/109)	seq1-1(x24)NB104215	NBBGa-1(x24)	1	NBBGa-1(27, 2, T)
	seq1-2(x15)NB118	NBBGa-2(x15)	2	NBBGa-2(16, 1)
	seq1-3(x5)NB211	NBBGa-3(x5)	3	NBBGa-3(5, T)
	seq1-4(x4)NB109209	NBBGa-4(x4)		NBBGa-4(4, 2)
	seq1-5(x2)NB124	NBBGa-5(x2)		NBBGa-5(2, 1)
	seq1-6(x2)NB228	NBBGa-6(x2)	1	
	seq1-7(x2)NB240	NBBGa-7(x2)	4	NBBGa-7(2, T)
	seq1-8(x2)NB216	NBBGa-8(x2)		NBBGa-8(2, 1)
	seq1-9(x2)NB217	NBBGa-9(x2)		NBBGa-9(2, 1)
	seq1-10(x1)NB105	NBBGa-10(x1)	2	
	seq1-11(x1)NB111	NBBGa-11(x1)		NBBGa-11(1, 1)
	seq1-12(x1)NB140	NBBGa-12(x1)	PCR error	
	seq1-13(x1)NB210	NBBGa-13(x1)	1	
	seq1-14(x1)NB225	NBBGa-14(x1)	5	NBBGa-14(2, T)
	seq1-15(x1)NB205	NBBGa-15(x1)	5	
	seq1-16(x1)NB201	NBBGa-16(x1)	PCR error	
	seq1-17(x1)NB219	NBBGa-17(x1)		NBBGa-17(1,1)
	seq1-18(x1)NB208	NBBGa-18(x1)		NBBGa-18(1,1)
	seq1-19(x1)NB229	NBBGa-19(x1)	PCR error	
	seq1-20(x1)NB236	NBBGa-20(x1)		NBBGa-20(1,1)
	seq1-21(x1)NB223	NBBGa-21(x1)		NBBGa-21(1,1)
	seq1-22(x27)NB320	NBBGa-22(x27)		NBBGa-22(27, V)
NBBGf(12/109)	seq2-1(x1)NB230	NBBGf-1(x1)		NBBGf-1(1,1)
	seq2-2(x1)NB214	NBBGf-2(x1)		NBBGf-2(1,1)
	seq2-3(x1)NB312	NBBGf-3(x10)		NBBGf-3(10, V)

(L2). All BG cDNA sequences found in T cells of line N (B21)

Line N (B21)	Representative Clones	Original Names	Note	Final Names
NTBGa(68/76)	seq1-1(x35)NT105201	NTBGa-1(x35)	1	NTBGa-1(45, 2, B)
	seq1-2(x8)NT234	NTBGa-2(x8)	1	
	seq1-3(x6)NT127235	NTBGa-3(x6)	4	NTBGa-3(6, B)
	seq1-4(x5)NT125	NTBGa-4(x5)	3	NTBGa-4(5, B)
	seq1-5(x2)NT111	NTBGa-5(x2)	PCR error	
	seq1-6(x2)NT104	NTBGa-6(x2)		NTBGa-6(2, 1)
	seq1-7(x1)NT101	NTBGa-7(x1)	1	
	seq1-8(x1)NT132	NTBGa-8(x1)	1	
	seq1-9(x1)NT108	NTBGa-9(x1)	PCR error	
	seq1-10(x1)NT208	NTBGa-10(x1)	5	NTBGa-10(2, B)
	seq1-11(x1)NT210	NTBGa-11(x1)		NTBGa-11(1,1)
	seq1-12(x1)NT212	NTBGa-12(x1)		NTBGa-12(1,1)
	seq1-13(x1)NT211	NTBGa-13(x1)		NTBGa-13(1,1)
	seq1-14(x1)NT221	NTBGa-14(x1)	PCR error	
	seq1-15(x1)NT239	NTBGa-15(x1)		NTBGa-15(1,1)
	seq1-16(x1)NT231	NTBGa-16(x1)	5	
NTBGb(4/76)	seq2(x4)NT237	NTBGb(x4)		NTBGb(4, 1)
NTBGc(2/76)	seq3(x2)NT110	NTBGc(x2)		NTBGc(2, 1)
NTBGd(1/76)	seq4(x1)NT138	NTBGd(x1)		NTBGd(1,1)
NTBGe(1/76)	seq6(x1)NT217	NTBGe(x1)		NTBGe(1,1)

(L3). All BG cDNA sequences found in B cells of line P2a (B19)

Line P2a (B19)	Representative Clones	Original Names	Note	Final Names
P2aBBGa(1/80)	seq1(x1)PB103	P2aBBGa(x1)		P2aBBGa(1, 1)
P2aBBGb(4/80)	seq2(x4)PB136	P2aBBGb(x4)	1	P2aBBGb(4, T)
P2aBBGc(75/80)	seq3-1(x2)PB147	P2aBBGc-1(x13)		P2aBBGc-1(13, 1st)
	seq3-2(x2)PB236	P2aBBGc-2(x13)	4	P2aBBGc-2(16, 1)
	seq3-3(x1)PB228	P2aBBGc-3(x8)		P2aBBGc-3(8, 2nd)
	seq3-4(x1)PB218	P2aBBGc-4(x8)	3	P2aBBGc-4(8, T)
	seq3-5(x1)PB108230	P2aBBGc-5(x2)		P2aBBGc-5(2, 2)
	seq3-6(x1)PB172	P2aBBGc-6(x2)		P2aBBGc-6(2, 1)
	seq3-7(x1)PB111	P2aBBGc-7(x1)		P2aBBGc-7(1, 1)
	seq3-8(x1)PB131	P2aBBGc-8(x1)		P2aBBGc-8(1, 1)
	seq3-9(x1)PB104	P2aBBGc-9(x1)		P2aBBGc-9(1, 1)
	seq3-10(x1)PB121	P2aBBGc-10(x1)		P2aBBGc-10(1, 1)
	seq3-11(x1)PB110	P2aBBGc-11(x1)	2	P2aBBGc-11(1, T)
	seq3-12(x1)PB107	P2aBBGc-12(x1)	PCR error	
	seq3-13(x1)PB232	P2aBBGc-13(x1)	4	
	seq3-14(x1)PB214	P2aBBGc-14(x1)	4	
	seq3-15(x1)PB233	P2aBBGc-15(x1)	4	
	seq3-16(x20)PB310	P2aBBGc-16(x20)		P2aBBGc-16(20, V)

(L4). All BG cDNA sequences found in T cells of line P2a (B19)

Line P2a (B19)	Representative Clones	Original Names	Note	Final Names
P2aTBGa(14/26)	seq1-1(x10)PT102221	P2aTBGa-1(x10)		P2aTBGa-1(10, 2)
	seq1-2(x3)PT122	P2aTBGa-2(x3)		P2aTBGa-2(3, 1)
	seq1-3(x1)PT202	P2aTBGa-3(x1)		P2aTBGa-3(1, 1)
P2aTBGb(6/26)	seq2-1(x2)PT104218	P2aTBGb-1(x2)	1	P2aTBGb-1(3, 2, B)
	seq2-2(x2)PT109	P2aTBGb-2(x2)		P2aTBGb-2(2, 1)
	seq2-3(x1)PT218	P2aTBGb-3(x1)	1	
	seq2-4(x1)PT125	P2aTBGb-4(x1)		P2aTBGb-4(1, 1)
P2aTBGc(6/26)	seq3-1(x1)PT117	P2aTBGc-1(x1)	2	P2aTBGc-1(1, B)
	seq3-2(x1)PT107	P2aTBGc-2(x1)		P2aTBGc-2(1, 1)
	seq3-3(x1)PT103	P2aTBGc-3(x1)		P2aTBGc-3(1, 1)
	seq3-4(x1)PT223	P2aTBGc-4(x1)		P2aTBGc-4(1, 1)
	seq3-5(x1)PT211	P2aTBGc-5(x1)	3	P2aTBGc-5(1, B)
	seq3-6(x1)PT228	P2aTBGc-6(x1)		P2aTBGc-6(1, 1)

(L5). All BG cDNA sequences found in B cells of line 6_1 (B2)

Line 61 (B2)	Representative Clones	New Names	Note	Final Names
6BBGa(5/63)	seq1-1(x3)L6B177	6BBGa-1(x3)		6BBGa-1(3, 1)
	seq1-2(x1)L6B190	6BBGa-2(x1)		6BBGa-2(1, 1)
	seq1-3(x1)L6B206	6BBGa-3(x1)		6BBGa-3(1, 1)
6BBGb(27/63)	seq2-1(x17)L6B181203	6BBGb-1(x17)	1	6BBGb-1(17, 2, T)
	seq2-2(x7)L6B189	6BBGb-2(x7)		6BBGb-2(7, 1)
	seq2-3(x3)L6B111	6BBGb-3(x3)		6BBGb-3(3, 1)
6BBGc(6/63)	seq3(x6)L6B158	6BBGc(x6)	2	6BBGc(6, T)
6BBGd(5/63)	seq4(x5)L6B218	6BBGd(x5)		6BBGd(5, 1)
6BBGf(12/63)	seq5-1(x7)L6B220	6BBGf-1(x7)		6BBGf-1(7, 1)
	seq5-2(x4)L6B166	6BBGf-2(x4)		6BBGf-2(4, 1)
	seq5-3(x1)L6B219	6BBGf-3(x1)		6BBGf-3(1, 1)
6BBGg(6/63)	seq6-1(x3)L6B152	6BBGg-1(x3)		6BBGg-1(3, 1)
	seq6-2(x3)L6B122	6BBGg-2(x3)		6BBGg-2(3, 1)
6BBGh(1/63)	seq7(x1)L6B211	6BBGh(x1)		6BBGh(1, 1)
6BBGi(1/63)	seq8(x1)L6B233	6BBGi(x1)		6BBGi(1, 1)

(L6). All BG cDNA sequences found in T cells of line 6_1 (B2)

Line 61 (B2)	Representative Clones	New Names	Note	Final Names
6TBGa(46/84)	seq1(x10)L6T108	6TBGa-1(x10)		6TBGa-1(10, 1)
	seq1-1(x9)L6T211	6TBGa-2(x9)		6TBGa-2(9, 1)
	seq1-2(x3)L6T224	6TBGa-3(x3)		6TBGa-3(3, 1)
	seq1-3(x2)L6T227	6TBGa-4(x2)		6TBGa-4(2, 1)
	seq1-4(x1)L6T222	6TBGa-5(x1)		6TBGa-5(1, 1)
	seq1(x21)L6T337	6TBGa-6(x21)		6TBGa-6(21, V)
6TBGb(15/84)	seq2(x7)L6T118	6TBGb-1(x7)	1	6TBGb-1(7, B)
	seq2-1(x6)L6T212	6TBGb-2(x6)		6TBGb-2(6, 1)
	seq2-2(x2)L6T221	6TBGb-3(x2)		6TBGb-3(2, 1)
6TBGc(11/84)	seq3-1(x6)L6T230	6TBGc-1(x6)	2	6TBGc-1(6, B)
	seq3-2(x3)L6T206	6TBGc-2(x3)		6TBGc-2(3, 1)
	seq3-3(x2)L6T228	6TBGc-3(x2)	PCR error	
6TBGd(10/84)	seq1(x10)L6T321	6TBGd(x10)		6TBGd(10, V)
6TBGe(2/84)	seq3(x2)L6T119	6TBGc-3(x2)		6TBGc-3(2, V)

(L7). All BG cDNA sequences found in B cells of line 15I (B15)

Line 15I (B15)	Representative Clones	Original Names	Note	Final Names
15iBBGa(35/46)	seq1-1(x8)L15B110209	15iBBGa-1(x8)	3	15iBBGa-1(10, 2, T)
	seq1-2(x5)L15B207	15iBBGa-2(x5)	1	15iBBGa-2(5, T)
	seq1-3(x3)L15B118227	15iBBGa-3(x3)	5	15iBBGa-3(4, 2, T)
	seq1-4(x3)L15B113	15iBBGa-4(x3)		15iBBGa-4(3, 1)
	seq1-5(x2)L15B120	15iBBGa-5(x2)		15iBBGa-5(2, 1)
	seq1-6(x2)L15B215	15iBBGa-6(x2)		15iBBGa-6(2, 1)
	seq1-7(x2)L15B218	15iBBGa-7(x2)	3	
	seq1-8(x1)L15B222	15iBBGa-8(x1)	5	
	seq1-9(x1)L15B124	15iBBGa-9(x1)	4	15iBBGa-9(1, T)
	seq1-10(x1)L15B123	15iBBGa-10(x1)		15iBBGa-10(1, 1)
	seq1-11(x1)L15B119	15iBBGa-11(x1)		15iBBGa-11(1, 1)
	seq1-12(x1)L15B127	15iBBGa-12(x1)		15iBBGa-12(1, 1)
	seq1-13(x1)L15B228	15iBBGa-13(x1)		15iBBGa-13(1, 1)
	seq1-14(x1)L15B216	15iBBGa-14(x1)		15iBBGa-14(1, 1)
	seq1-15(x1)L15B220	15iBBGa-15(x1)	PCR error	
	seq1-16(x1)L15B240	15iBBGa-16(x1)	2	15iBBGa-16(1, T)
	seq1-17(x1)L15B232	15iBBGa-17(x1)	5'UTR var	15iBBGa-17(1, 1)
15iBBGb(7/46)	seq2-1(x2)L15B125214	15iBBGb-1(x2)	6	15iBBGb-1(3, 2, T)
	seq2-2(x2)L15B205	15iBBGb-2(x2)	7	15iBBGb-2(3, 2, T)
	seq2-3(x1)L15B128	15iBBGb-3(x1)	6	
	seq2-4(x1)L15B112	15iBBGb-4(x1)	7	
	seq2-5(x1)L15B231	15iBBGb-5(x1)	PCR error	
15iBBGc(2/46)	seq3(x2)L15B229	15iBBGc(x2)		15iBBGc(2, 1)
15iBBGd(1/46)	seq4(x1)L15B130	15iBBGd(x1)		15iBBGd(1, 1)
15iBBGe(1/46)	seq5(x1)L15B135	15iBBGe(x1)		15iBBGe(1, 1)

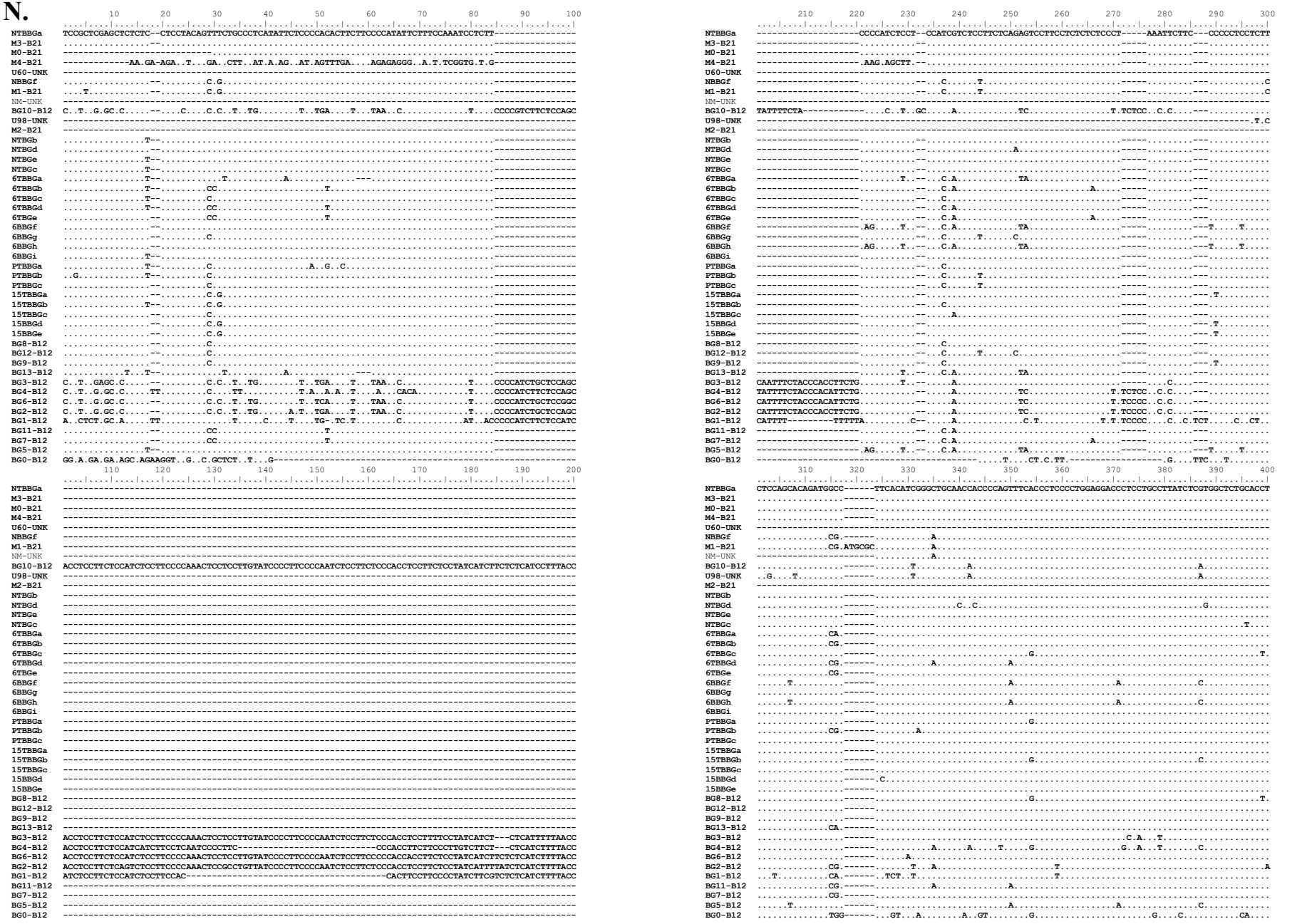
(L8). All BG cDNA sequences found in T cells of line 15I (B15)

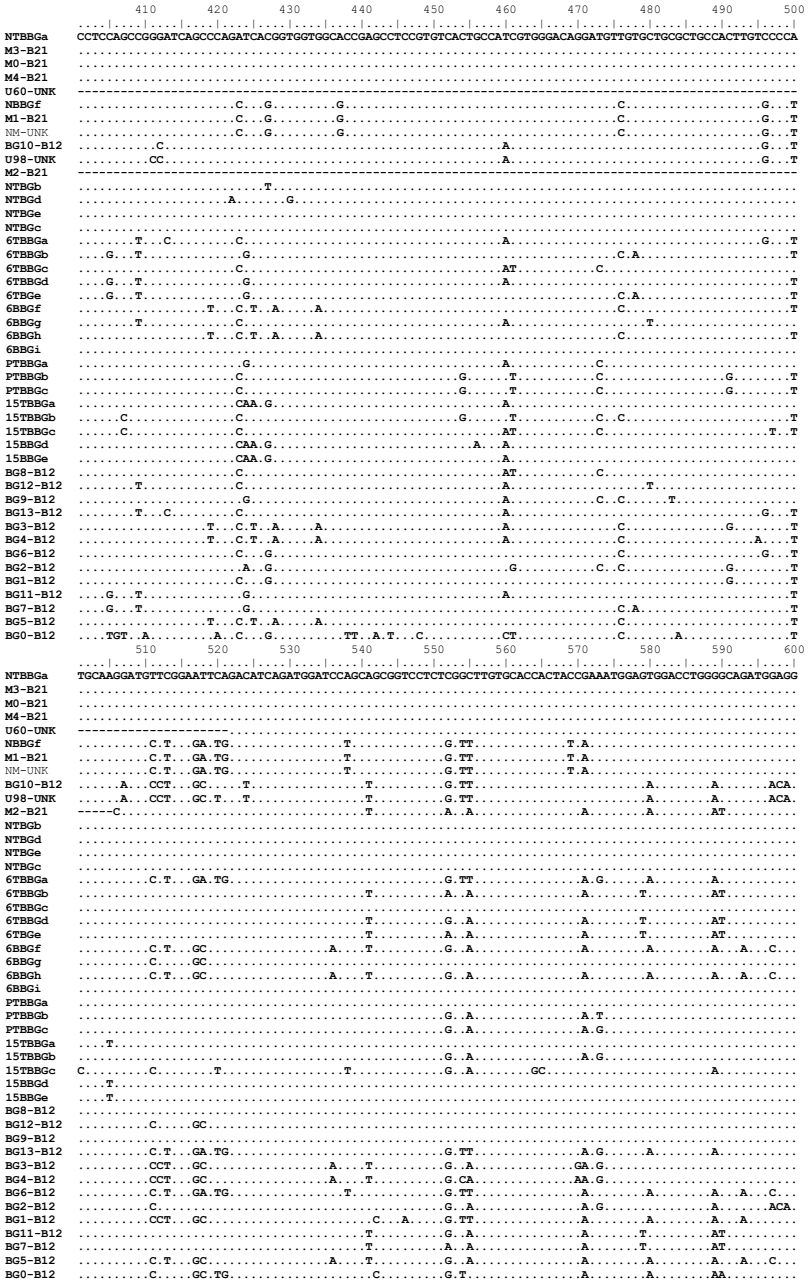
Line 15I (B15)	Representative Clones	Original Names	Note	Final Names
15iTBGa(24/39)	seq1-1(x6)L15T239	15iTBGa-1(x6)	3	15iTBGa-1(6, B)
	seq1-2(x4)L15T127208	15iTBGa-2(x4)	4	15iTBGa-2(4, 2, B)
	seq1-3(x3)L15T118216	15iTBGa-3(x3)		15iTBGa-3(3, 2)
	seq1-4(x3)L15T102	15iTBGa-4(x3)	5	15iTBGa-4(3, B)
	seq1-5(x2)L15T117	15iTBGa-5(x2)	2	15iTBGa-5(2, B)
	seq1-6(x2)L15T226	15iTBGa-6(x2)		15iTBGa-6(2, 1)
	seq1-7(x1)L15T140	15iTBGa-7(x1)	1	15iTBGa-7(1, B)
	seq1-8(x1)L15T105	15iTBGa-8(x1)		15iTBGa-8(1, 1)
	seq1-9(x1)L15T218	15iTBGa-9(x1)	5'UTR var	15iTBGa-9(1, 1)
	seq1-10(x1)L15T134	15iTBGa-10(x1)	PCR error	
15iTBGb(11/39)	seq2-1(x4)L15T126207	15iTBGb-1(x4)	6	15iTBGb-1(4, 2, B)
	seq2-2(x2)L15T128215	15iTBGb-2(x2)		15iTBGb-2(2, 2)
	seq2-3(x1)L15T123	15iTBGb-3(x1)	PCR error	
	seq2-4(x1)L15T209	15iTBGb-4(x1)		15iTBGb-4(1, 1)
	seq2-5(x1)L15T219	15iTBGb-5(x1)	7	15iTBGb-5(1, B)
	seq2-6(x1)L15T227	15iTBGb-6(x1)	PCR error	
	seq2-7(x1)L15T237	15iTBGb-7(x1)		15iTBGb-7(1, 1)
15iTBGc(4/39)	seq3-1(x3)L15T201	15iTBGc-1(x3)		15iTBGc-1(3, 1)
	seq3-2(x1)L15T210	15iTBGc-2(x1)	PCR error	

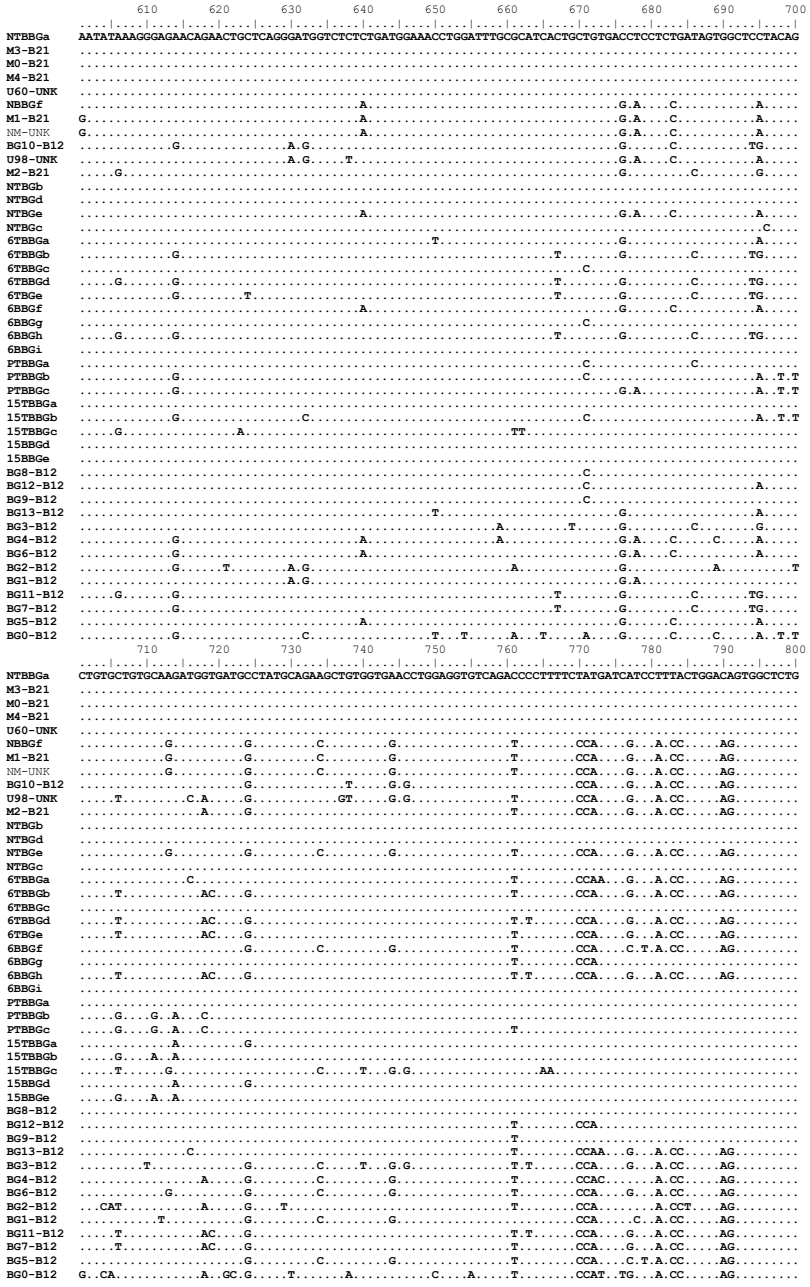
Appendix M. Results for BLAT chicken genomic sequence (Gallus gullas 5.0) using nucleotide sequence of exon 2 of BG8 gene (B12 haplotype)

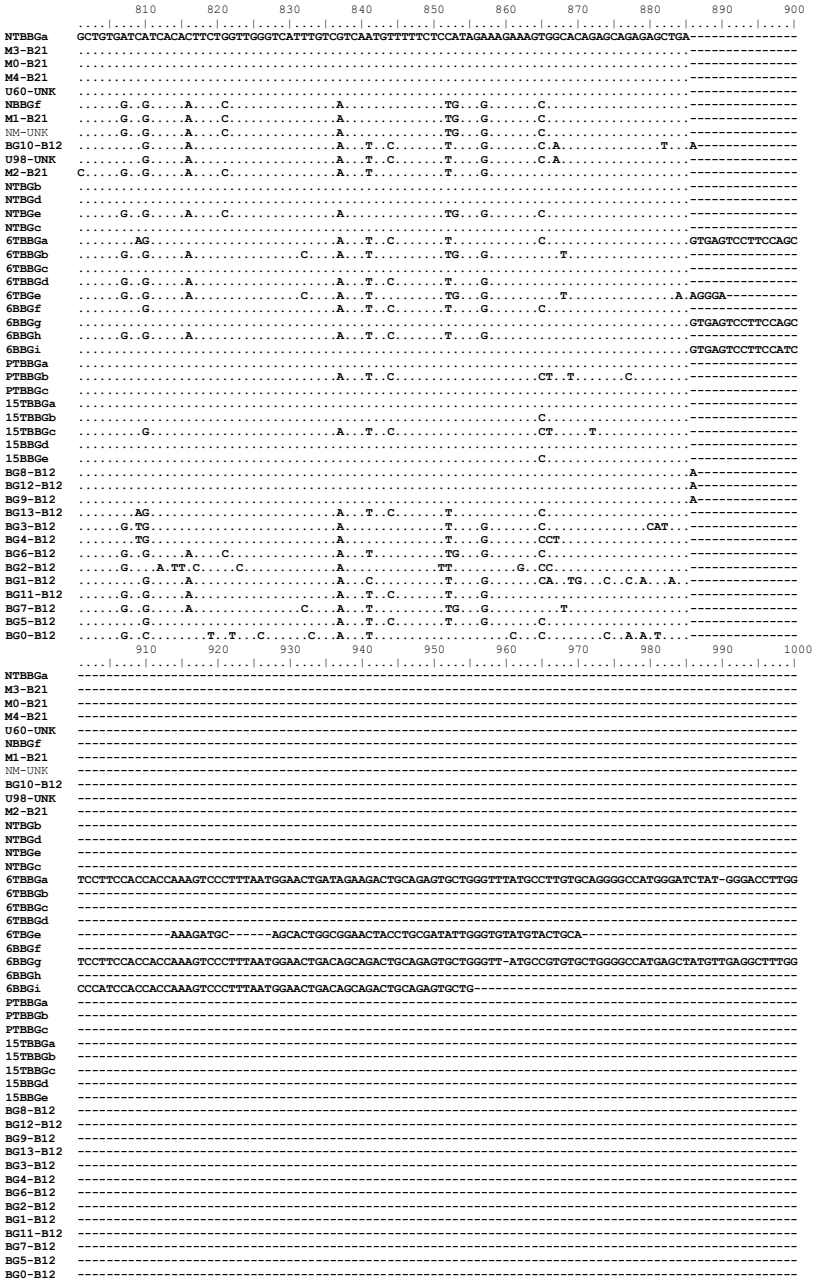
Genomic Location	Overlapping Gene(s)	Orientation	Query start	Query end	Length	Score	E-val	%ID
16:293702-294043 [Sequence]	KIFC1, ENSGALG00000040101, ENSGALG00000040185, ENSGALG00000040179, ENSGALG00000041588	Reverse	1	342	342 [Sequence]	654.0	5.1e-188	97.95 [Alignment]
16:378125-378466 [Sequence]	ENSGALG00000024357	Reverse	1	342	342 [Sequence]	604.0	5.7e-173	93.86 [Alignment]
16:317691-318032 [Sequence]	ENSGALG00000040101, ENSGALG00000041588, ENSGALG00000032221	Reverse	1	342	342 [Sequence]	595.0	3.0e-170	93.27 [Alignment]
16:368271-368612 [Sequence]	ENSGALG00000040101, ENSGALG00000041588, ENSGALG00000044585, ENSGALG00000024357	Reverse	1	342	342 [Sequence]	594.0	4.3e-170	93.86 [Alignment]
16:372049-372390 [Sequence]	ENSGALG00000040101, ENSGALG00000041588, ENSGALG00000044585, ENSGALG00000024357	Reverse	1	342	342 [Sequence]	594.0	4.7e-170	93.57 [Alignment]
16:278737-279078 [Sequence]	KIFC1, ENSGALG00000040101, ENSGALG00000040185, ENSGALG00000040179, ENSGALG00000039480	Reverse	1	342	342 [Sequence]	584.0	4.7e-167	92.69 [Alignment]
16:271141-271482 [Sequence]	KIFC1, ENSGALG00000040101, ENSGALG00000040185, ENSGALG00000040179, ENSGALG00000045233	Reverse	1	342	342 [Sequence]	584.0	5.9e-167	91.81 [Alignment]
16:347711-348052 [Sequence]	ENSGALG00000040101, ENSGALG00000041588, ENSGALG00000044585, ENSGALG00000024357	Reverse	1	342	342 [Sequence]	583.0	1.1e-166	92.69 [Alignment]
16:361155-361496 [Sequence]	ENSGALG00000040101, ENSGALG00000041588, ENSGALG00000044585, ENSGALG00000024357, ENSGALG00000024365	Reverse	1	342	342 [Sequence]	583.0	1.2e-166	92.69 [Alignment]
16:129517-129858 [Sequence]	BG1	Forward	1	342	342 [Sequence]	581.0	2.8e-166	92.40 [Alignment]
KQ759463.1:10986-11327 [Sequence]	ENSGALG00000044751	Forward	1	342	342 [Sequence]	581.0	2.8e-166	92.40 [Alignment]
KQ759463.1:517-858 [Sequence]	ENSGALG00000045381	Reverse	1	342	342 [Sequence]	581.0	2.8e-166	92.40 [Alignment]
16:333719-334060 [Sequence]	ENSGALG00000040101, ENSGALG00000041588, ENSGALG00000044585	Reverse	1	342	342 [Sequence]	577.0	4.7e-165	91.52 [Alignment]
16:354467-354808 [Sequence]	ENSGALG00000040101, ENSGALG00000041588, ENSGALG00000044585, ENSGALG00000024357, ENSGALG00000028157	Reverse	1	342	342 [Sequence]	577.0	6.6e-165	91.81 [Alignment]
16:261619-261960 [Sequence]	KIFC1, ENSGALG00000040101, ENSGALG00000040185, ENSGALG00000040179, ENSGALG00000044802	Reverse	1	342	342 [Sequence]	575.0	3.1e-164	91.23 [Alignment]
16:324768-325108 [Sequence]	ENSGALG00000040101, ENSGALG00000041588, ENSGALG00000044585	Reverse	2	342	341 [Sequence]	558.0	4.1e-159	90.91 [Alignment]
16:340893-341239 [Sequence]	ENSGALG00000040101, ENSGALG00000041588, ENSGALG00000044585, ENSGALG00000024357, ENSGALG00000045523	Reverse	1	342	347 [Sequence]	544.0	5.7e-155	89.05 [Alignment]
16:285985-286297 [Sequence]	KIFC1, ENSGALG00000040101, ENSGALG00000040185, ENSGALG00000040179, ENSGALG00000041588, ENSGALG00000044555	Reverse	1	313	313 [Sequence]	534.0	6.2e-152	92.33 [Alignment]
2:98306686-98307027 [Sequence]	ENSGALG00000022875	Reverse	1	342	342 [Sequence]	516.0	1.2e-146	85.96 [Alignment]
16:252272-252570 [Sequence]	KIFC1, ENSGALG00000040101	Reverse	1	299	299 [Sequence]	492.0	2.4e-139	89.97 [Alignment]
AADN04011814.1:5727-5968 [Sequence]	ENSGALG00000040485	Reverse	101	342	242 [Sequence]	399.0	2.7e-111	91.74 [Alignment]
KQ759463.1:19241-19443 [Sequence]		Forward	140	342	203 [Sequence]	365.0	3.7e-101	95.07 [Alignment]
KQ759463.1:19101-19231 [Sequence]		Forward	1	131	131 [Sequence]	225.0	7.7e-59	93.13 [Alignment]
AADN04011814.1:5976-6069 [Sequence]	ENSGALG00000040485	Reverse	1	94	94 [Sequence]	172.0	4.2e-43	95.74 [Alignment]
16:252226-252254 [Sequence]	KIFC1, ENSGALG00000040101	Reverse	314	342	29 [Sequence]	46.0	5.8e-05	89.66 [Alignment]

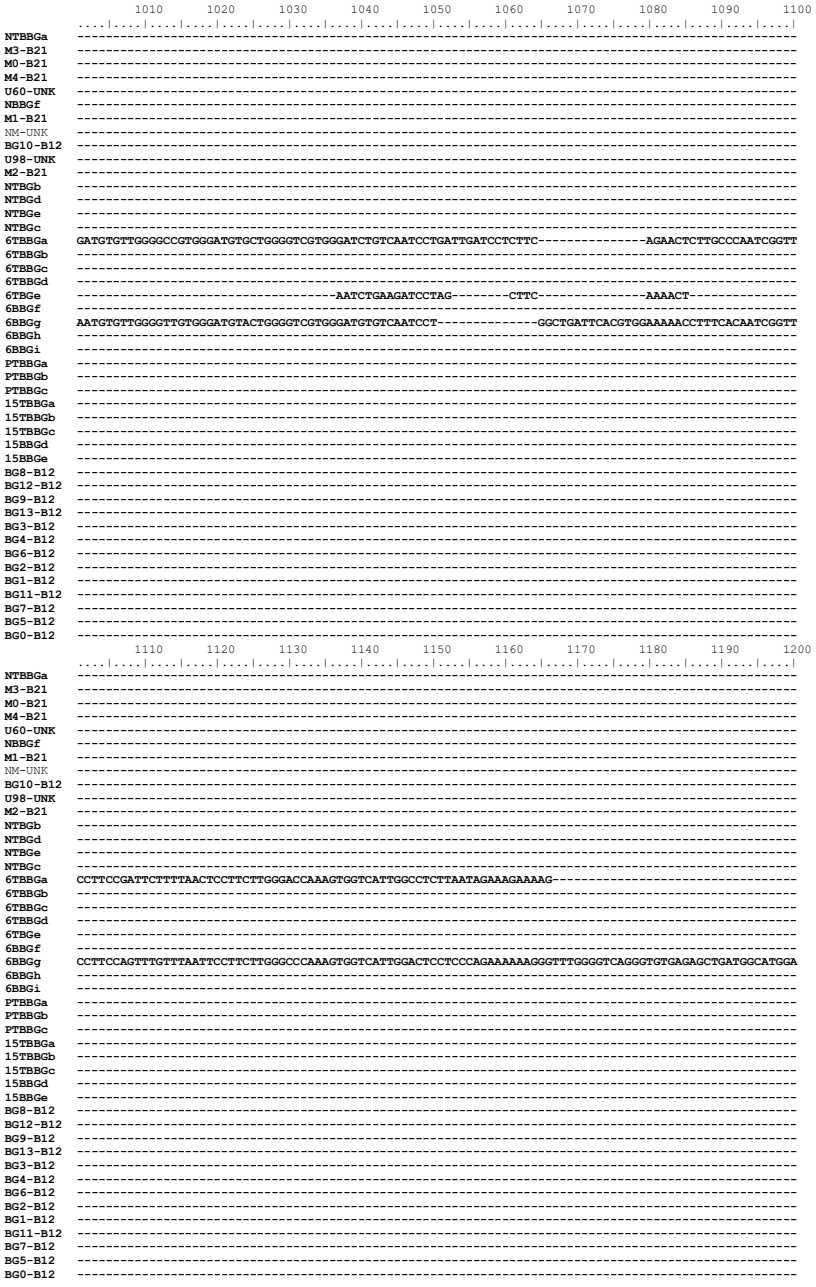
Appendix N.

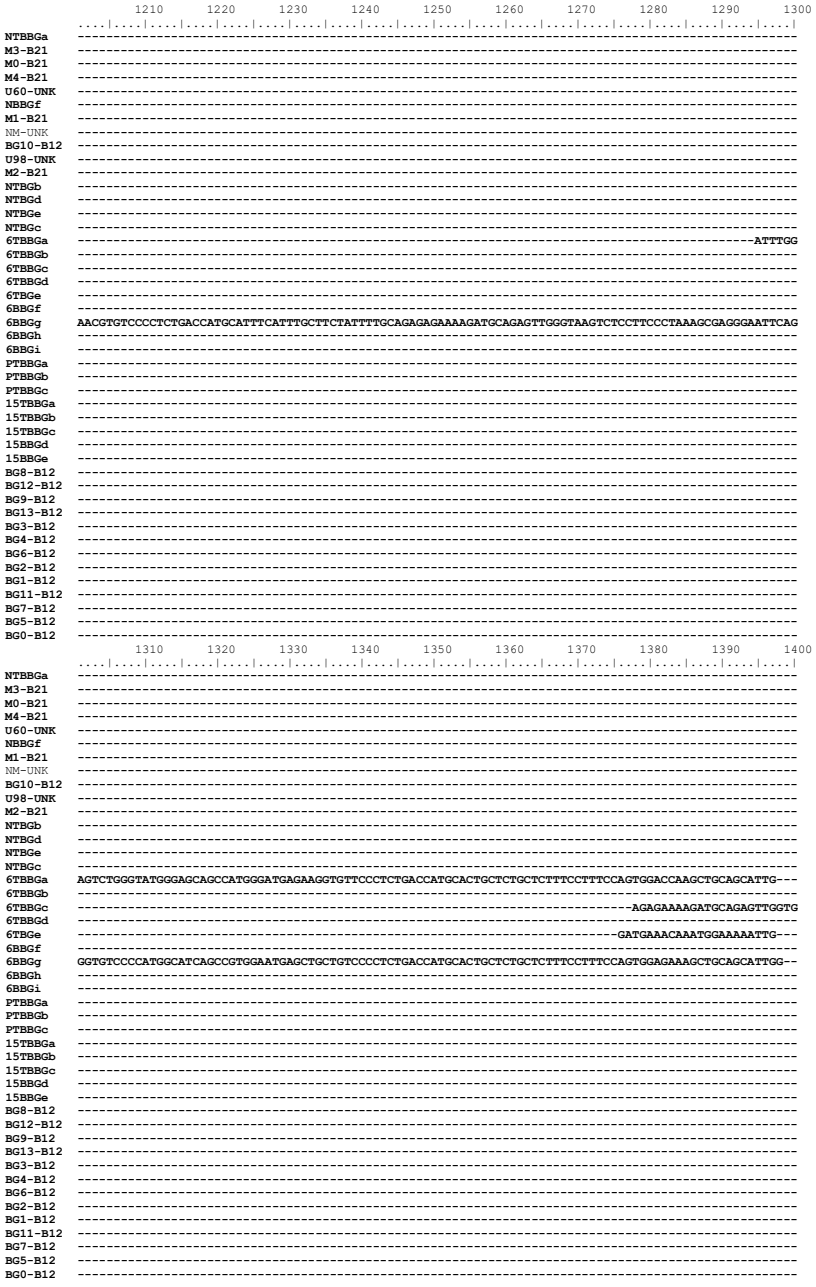


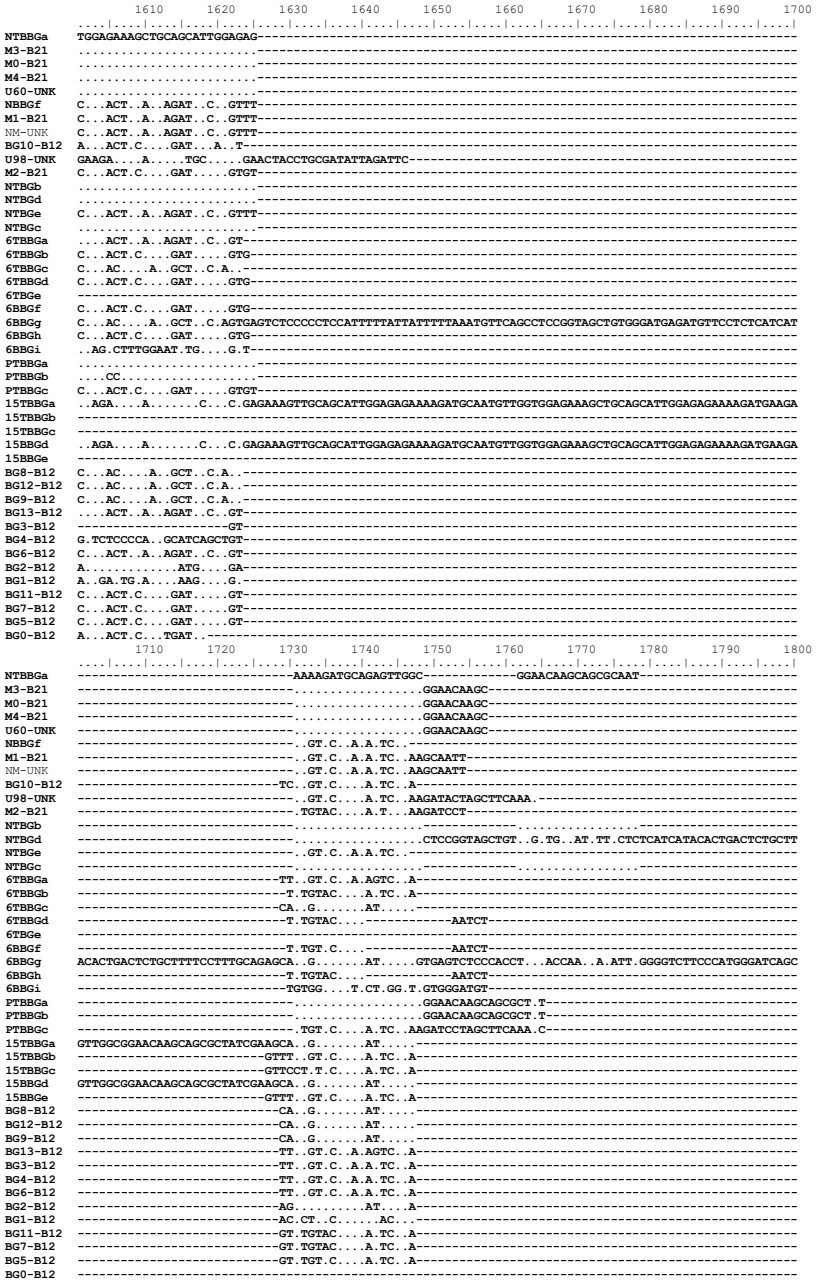

410 420 430 440 450 460 470 480 490 500
NTBBGa CCTCCAGCCGGGATCAGCCCAGATCACGGTGGTGGCACCGAGCCTCCGTGTCACTGCCATCGTGGGACAGGATGTTGTGCTGCGCTGCCACTTGTCCCCA
M3-B21
M0-B21
M4-B21
U60-UNK
NBBGf
M1-B21
NM-UNK
BG10-B12
U98-UNK
M2-B21
NTBGb
NTBGd
NTBGe
NTBGc
6TBBGa
6TBBGb
6TBBGc
6TBBGd
6TBGe
6BBGf
6BBGg
6BBGh
6BBGi
PTBBGa
PTBBGb
PTBBGc
15TBBGa
15TBBGb
15TBBGc
15BBGd
15BBGe
BG8-B12
BG12-B12
BG9-B12
BG13-B12
BG3-B12
BG4-B12
BG6-B12
BG2-B12
BG1-B12
BG11-B12
BG7-B12
BG5-B12
BG0-B12
510 520 530 540 550 560 570 580 590 600
NTBBGa TGCAAGGATGTTCGGAATTCAGACATCAGATGGATCCAGCAGCGGTCCTCTCGGCTTGTGCACCACTACCGAAATGGAGTGGACCTGGGGCAGATGGAGG

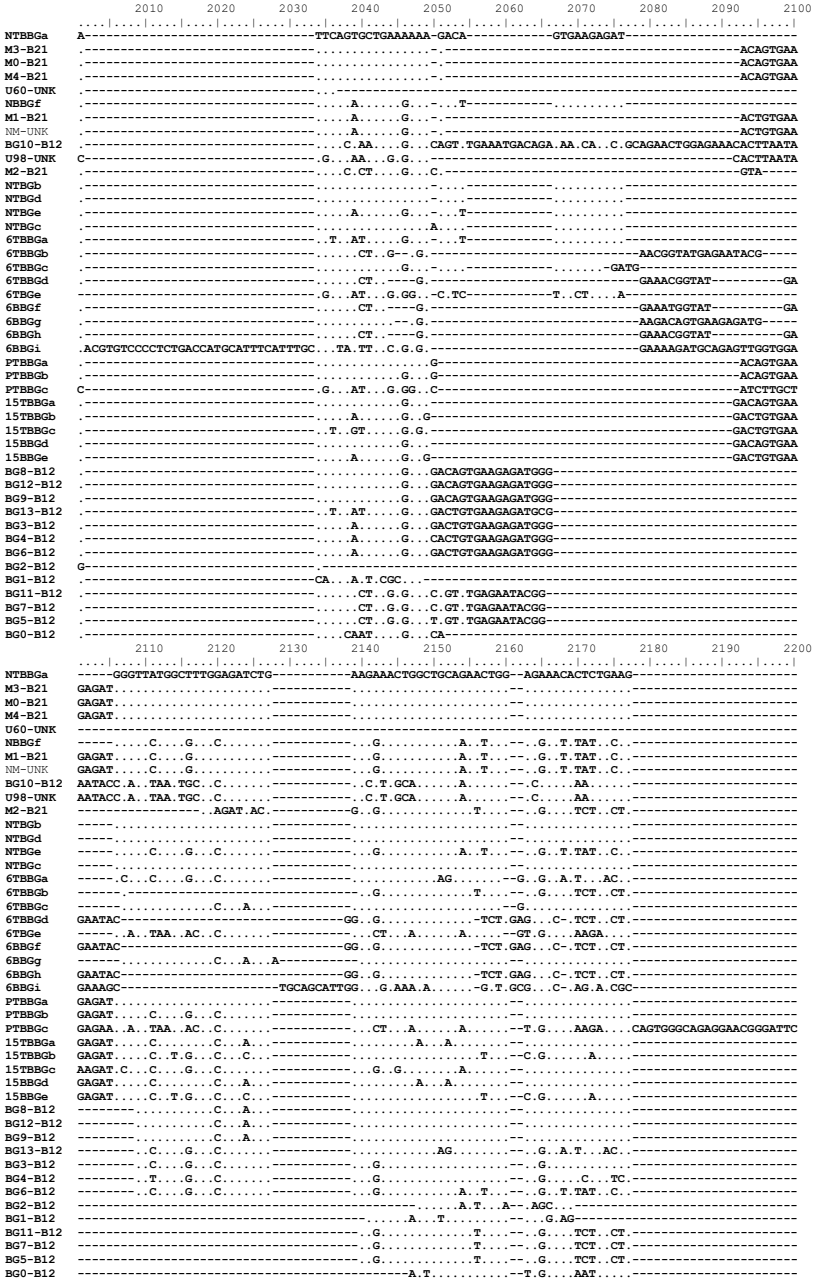

610 620 630 640 650 660 670 680 690 700
NTBBGa AATATAAAGGGAGAACAGAACTGCTCAGGGATGGTCTCTCTGATGGAAACCTGGATTTGCGCATCACTGCTGTGACCTCCTCTGATAGTGGCTCCTACAG
M3-B21
M0-B21
M4-B21
U60-UNK
NBBGf
M1-B21
NM-UNK
BG10-B12
U98-UNK
M2-B21
NTBGb
NTBGd
NTBGe
NTBGc
6TBBGa
6TBBGb
6TBBGc
6TBBGd
6TBGe
6BBGf
6BBGg
6BBGh
6BBGi
PTBBGa
PTBBGb
PTBBGc
15TBBGa
15TBBGb
15TBBGc
15BBGd
15BBGe
BG8-B12
BG12-B12
BG9-B12
BG13-B12
BG3-B12
BG4-B12
BG6-B12
BG2-B12
BG1-B12
BG11-B12
BG7-B12
BG5-B12
BG0-B12
710 720 730 740 750 760 770 780 790 800
NTBBGa CTGTGCTGTGCAAGATGGTGATGCCTATGCAGAAGCTGTGGTGAACCTGGAGGTGTCAGACCCCTTTTCTATGATCATCCTTTACTGGACAGTGGCTCTG

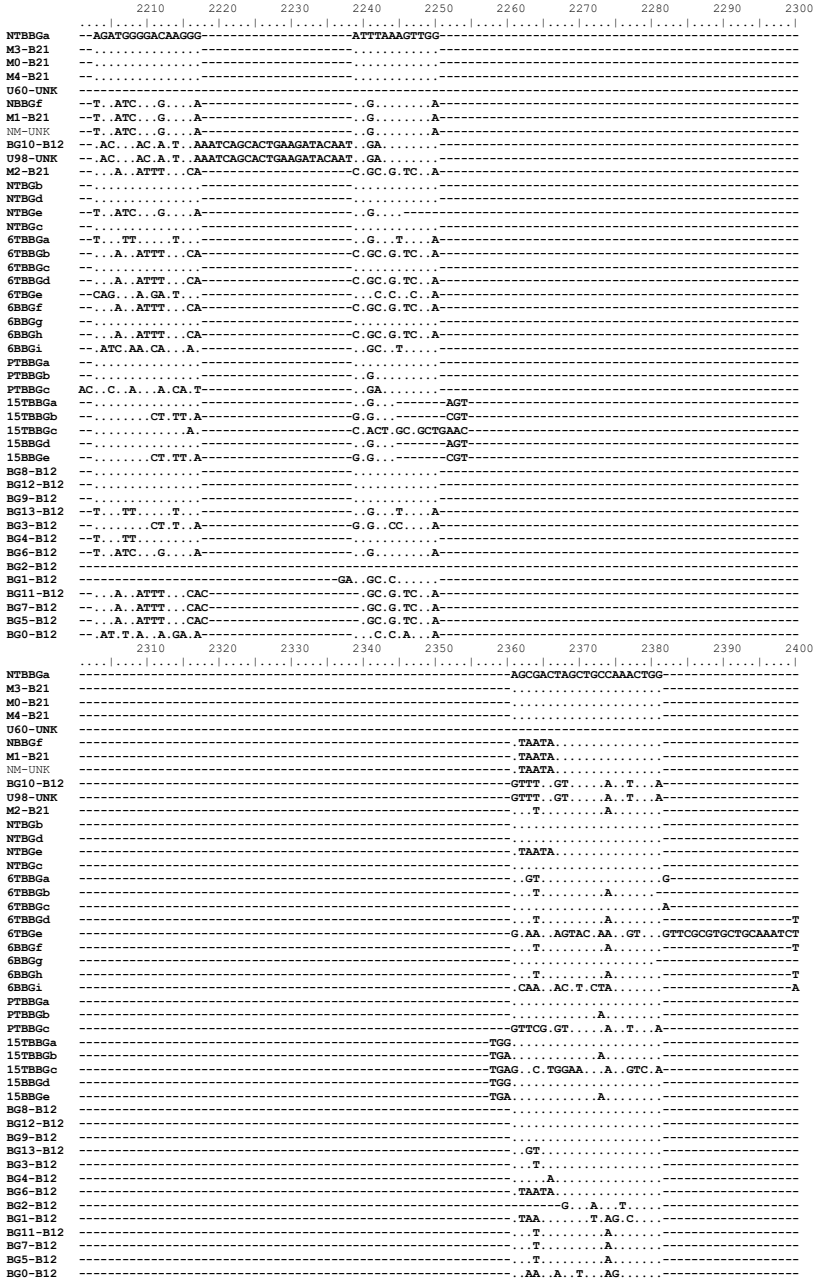

NTBBGa
M3-B21
M0-B21
M4-B21
U60-UNK
NBBGf
M1-B21
NM-UNK
BG10-B12
U98-UNK
M2-B21
NTBGb
NTBGd
NTBGe
NTBGc
6TBBGa
6TBBGb
6TBBGc
6TBBGd
6TBGe
6BBGf
6BBGg
6BBGh
6BBGi
PTBBGa
PTBBGb
PTBBGc
15TBBGa
15TBBGb
15TBBGc
15BBGd
15BBGe
BG8-B12
BG12-B12
BG9-B12
BG13-B12
BG3-B12
BG4-B12
BG6-B12
BG2-B12
BG1-B12
BG11-B12
BG7-B12
BG5-B12
BG0-B12

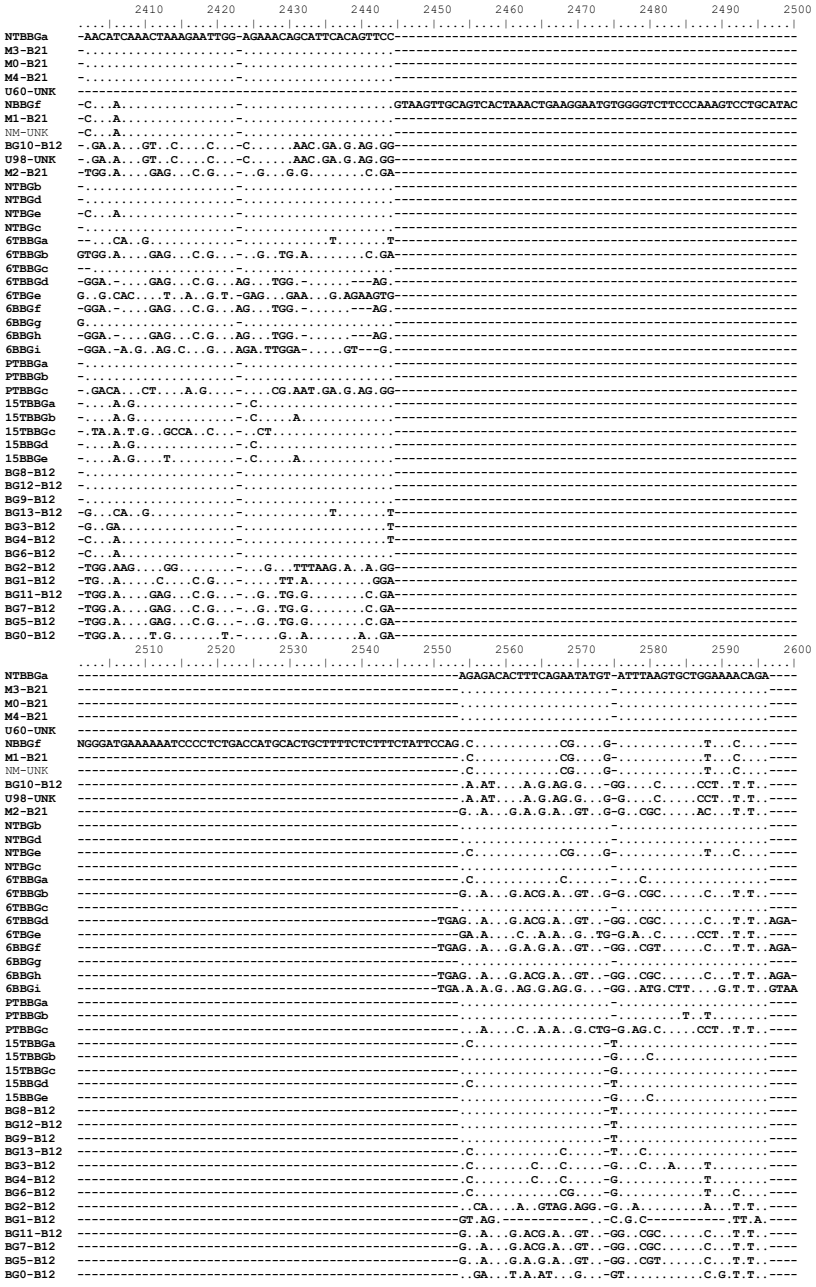

NTBBGa
M3-B21
M0-B21
M4-B21
U60-UNK
NBBGf
M1-B21
NM-UNK
BG10-B12
U98-UNK
M2-B21
NTBGb
NTBGd
NTBGe
NTBGc
6TBBGa
6TBBGb
6TBBGc
6TBBGd
6TBGe
6BBGf
6BBGg
6BBGh
6BBGi
PTBBGa
PTBBGb
PTBBGc
15TBBGa
15TBBGb
15TBBGc
15BBGd
15BBGe
BG8-B12
BG12-B12
BG9-B12
BG13-B12
BG3-B12
BG4-B12
BG6-B12
BG2-B12
BG1-B12
BG11-B12
BG7-B12
BG5-B12
BG0-B12

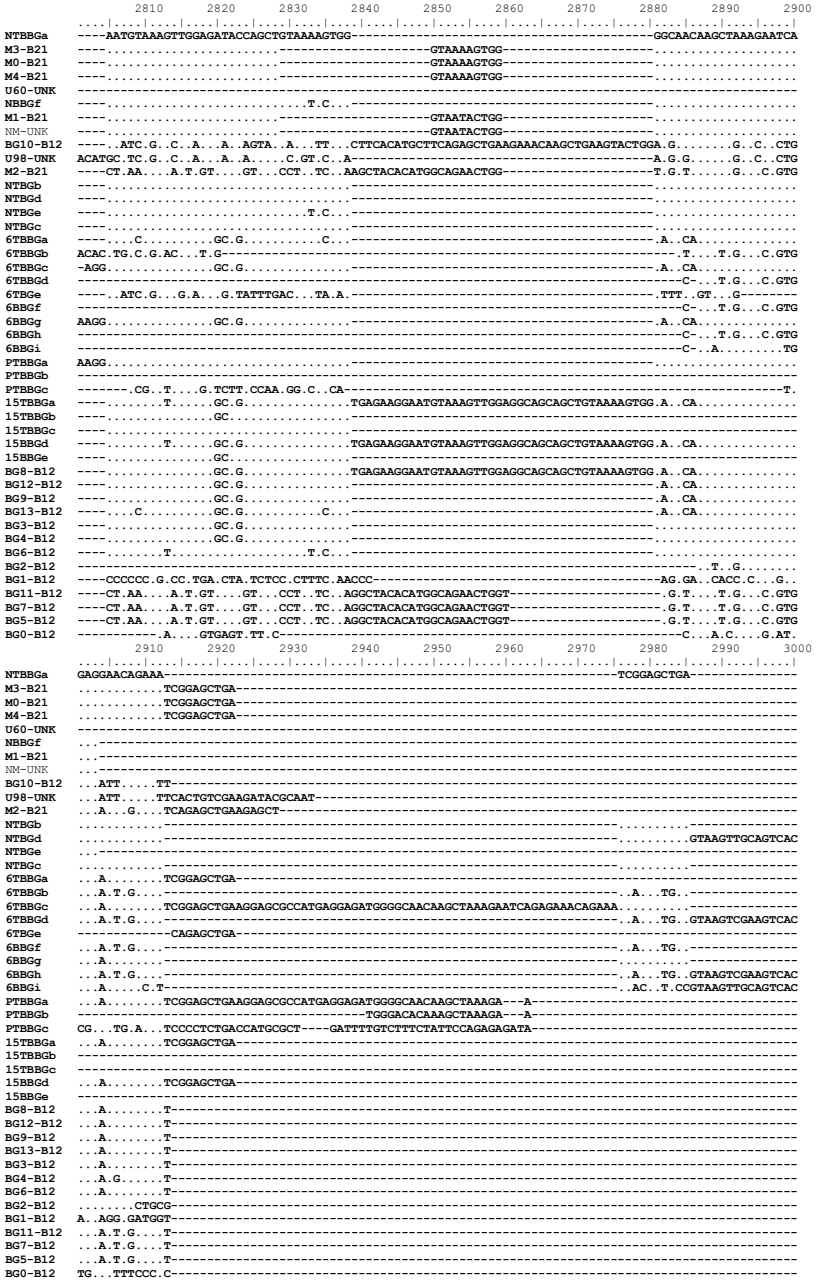

1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
NTBBGa
M3-B21
M0-B21
M4-B21
U60-UNK
NBBGf
M1-B21
NM-UNK
BG10-B12
U98-UNK
M2-B21
NTBGb
NTBGd
NTBGe
NTBGc
6TBBGa ATTTGG
6TBBGb
6TBBGc
6TBBGd
6TBGe
6BBGf
6BBGg AACGTGTCCCCTCTGACCATGCATTTCATTTGCTTCTATTTTGCAGAGAGAAAAGATGCAGAGTTGGGTAAGTCTCCTTCCCTAAAGCGAGGGAATTCAG
6BBGh
6BBGi
PTBBGa
PTBBGb
PTBBGc
15TBBGa
15TBBGb
15TBBGc
15BBGd
15BBGe
BG8-B12
BG12-B12
BG9-B12
BG13-B12
BG3-B12
BG4-B12
BG6-B12
BG2-B12
BG1-B12
BG11-B12
BG7-B12
BG5-B12
BG0-B12
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
NTBBGa
M3-B21
M0-B21
M4-B21
U60-UNK
NBBGf
M1-B21
NM-UNK
BG10-B12
U98-UNK
M2-B21
NTBGb
NTBGd
NTBGe
NTBGc
6TBBGa AGTCTGGGTATGGGAGCAGCCATGGGATGAGAAGGTGTTCCCTCTGACCATGCACTGCTCTGCTCTTTCCTTTCCAGTGGACCAAGCTGCAGCATTG---
6TBBGb
6TBBGc AGAGAAAAGATGCAGAGTTGGTG
6TBBGd
6TBGe GATGAAACAAATGGAAAAATTG---
6BBGf
6BBGg GGTGTCCCCATGGCATCAGCCGTGGAATGAGCTGCTGTCCCCTCTGACCATGCACTGCTCTGCTCTTTCCTTTCCAGTGGAGAAAGCTGCAGCATTGG--
6BBGh
6BBGi
PTBBGa
PTBBGb
PTBBGc
15TBBGa
15TBBGb
15TBBGc
15BBGd
15BBGe
BG8-B12
BG12-B12
BG9-B12
BG13-B12
BG3-B12
BG4-B12
BG6-B12
BG2-B12
BG1-B12
BG11-B12
BG7-B12
BG5-B12
BG0-B12
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
NTBBGa AGAGAAAAAGATGC AGAGTTGG
M3-B21 .....GTAAGTCTC-
M0-B21
M4-B21
U60-UNK
NBBGf ..C.....
M1-B21 ..C.....
NM-UNK ..C.....
BG10-B12 AAAAAAAGATGCAATGT---TGGATCAAGTTGTGACAATGAGA ..T.C...
U98-UNK .A.....AT.....
M2-B21 ..G.....CAC...
NTBGb
NTBGd
NTBGe ..C.....
NTBGc
6TBBGa G.....GTAAGTCTCC
6TBBGb ..G.....CAC...
6TBBGc GAGAAAGCTGCAGCAT--TGG.....
6TBBGd ..G.....CAC...
6TBGe G...TTC.GA..T.....TC....A.....
6BBGf ..G.G.....CAC...
6BBGg .....GTAAGTCTCC
6BBGh ..G.....CAC...
6BBGi G.TT.TGCC...TGCTGGGGCCATG..CTA..T.....
PTBBGa
PTBBGb
PTBBGc ..G.....CAC...
15TBBGa
15TBBGb .....AT.....
15TBBGc ..C.....
15BBGd
15BBGe .....AT.....
BG8-B12 ----GAGAAAAGATGCAGAGTTGGTGGAGAAAGCTGCAGCATTGG.....
BG12-B12 ----GAGAAAAGATGCAGAGTTGGTGGAGAAAGCTGCAGCATTGG.....
BG9-B12 ----GAGAAAAGATGCAGAGTTGGTGGAGAAAGCTGCAGCATTGG.....
BG13-B12 ----TGGACCAAGCTGCAGCATTGG.....
BG3-B12 .....AT.....
BG4-B12 .....AT.....
BG6-B12 .....C.....
BG2-B12 .....CA....
BG1-B12 ..G.....
BG11-B12 ..G.....CAC...
BG7-B12 ..G.....CAC...
BG5-B12 ..G.G.....CAC...
BG0-B12 ...T.....T.....
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
NTBBGa
M3-B21
M0-B21
M4-B21
U60-UNK
NBBGf
M1-B21
NM-UNK
BG10-B12
U98-UNK
M2-B21
NTBGb
NTBGd
NTBGe
NTBGc
6TBBGa TTCCCTAAAGCGAGGGAATTCAGGGTCTCCCCATGGCATCAGCTGTGGGATGAGCAGCTGTCCTCTCTGACCATGCACTGCTCTGCTCTTTCTTTTCCAG
6TBBGb
6TBBGc
6TBBGd
6TBGe
6BBGf
6BBGg TTCCCCACAGCGAGGGAATTCAGGGTTTCCCCATGGCGTTAGCCACGGGATGAG-CAGCTGTCCTCTCTGACCATGCACTGCTCTGCTCTTTCTTTCCAG
6BBGh
6BBGi
PTBBGa
PTBBGb
PTBBGc
15TBBGa TGGAGAAAGCTGCAGCATTGG
15TBBGb
15TBBGc
15BBGd TGGAGAAAGCTGCAGCATTGG
15BBGe
BG8-B12
BG12-B12
BG9-B12
BG13-B12
BG3-B12
BG4-B12
BG6-B12
BG2-B12
BG1-B12
BG11-B12
BG7-B12
BG5-B12
BG0-B12

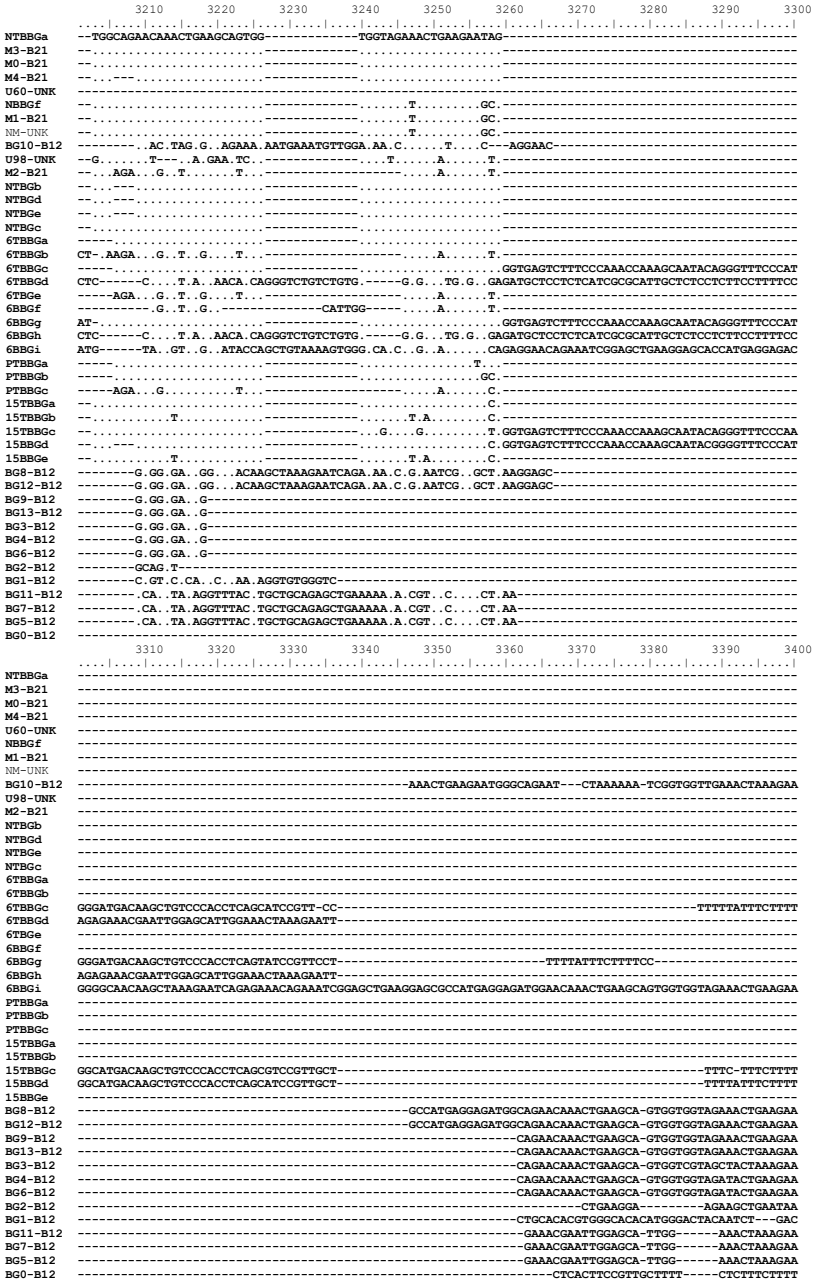

1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
NTBBGa TGGAGAAAGCTGCAGCATTGGAGAG
M3-B21
M0-B21
M4-B21
U60-UNK
NBBGf C...ACT..A..AGAT..C..GTTT
M1-B21 C...ACT..A..AGAT..C..GTTT
NM-UNK C...ACT..A..AGAT..C..GTTT
BG10-B12 A...ACT.C....GAT...A..T
U98-UNK GAAGA...A.....TGC....GAACTACCTGCGATATTAGATTC
M2-B21 C...ACT.C....GAT.....GTGT
NTBGb
NTBGd
NTBGe C...ACT..A..AGAT..C..GTTT
NTBGc
6TBBGa ....ACT..A..AGAT..C..GT
6TBBGb C...ACT.C....GAT.....GTG
6TBBGc C...AC....A..GCT..C.A..
6TBBGd C...ACT.C....GAT.....GTG
6TBGe
6BBGf C...ACT.C....GAT.....GTG
6BBGg C...AC....A..GCT..C.AGTGAGTCTCCCCCTCCATTTTTATTATTTTTAAATGTTCAGCCTCCGGTAGCTGTGGGATGAGATGTTCCTCTCATCAT
6BBGh C...ACT.C....GAT.....GTG
6BBGi ..AG.CTTTGGAAT.TG....G.T
PTBBGa
PTBBGb ....CC.....
PTBBGc C...ACT.C....GAT.....GTGT
15TBBGa ..AGA....A.......C...C.GAGAAAGTTGCAGCATTGGAGAGAAAAGATGCAATGTTGGTGGAGAAAGCTGCAGCATTGGAGAGAAAAGATGAAGA
15TBBGb
15TBBGc
15BBGd ..AGA....A.......C...C.GAGAAAGTTGCAGCATTGGAGAGAAAAGATGCAATGTTGGTGGAGAAAGCTGCAGCATTGGAGAGAAAAGATGAAGA
15BBGe
BG8-B12 C...AC....A..GCT..C.A..
BG12-B12 C...AC....A..GCT..C.A..
BG9-B12 C...AC....A..GCT..C.A..
BG13-B12 ....ACT..A..AGAT..C..GT
BG3-B12 ----------GT
BG4-B12 G.TCTCCCCA..GCATCAGCTGT
BG6-B12 C...ACT..A..AGAT..C..GT
BG2-B12 A.............ATG....GA
BG1-B12 A..GA.TG.A....AAG....G..
BG11-B12 C...ACT.C....GAT.....GT
BG7-B12 C...ACT.C....GAT.....GT
BG5-B12 C...ACT.C....GAT.....GT
BG0-B12 A...ACT.C...TGAT..
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
NTBBGa AAAAGATGCAGAGTTGGC GGAACAAGCAGCGCAAT
M3-B21 GGAACAAGC
M0-B21 GGAACAAGC
M4-B21 GGAACAAGC
U60-UNK GGAACAAGC
NBBGf GT.C..A.A.TC..
M1-B21 GT.C..A.A.TC..AAGCAATT
NM-UNK GT.C..A.A.TC..AAGCAATT
BG10-B12 TC..GT.C....A.TC..A
U98-UNK GT.C....A.TC..AAGATACTAGCTTCAAA.
M2-B21 TGTAC....A.T...AAGATCCT
NTBGb
NTBGd CTCCGGTAGCTGT..G.TG..AT.TT.CTCTCATCATACACTGACTCTGCTT
NTBGe GT.C..A.A.TC..
NTBGc
6TBBGa TT..GT.C..A.AGTC..A
6TBBGb T.TGTAC....A.TC..A
6TBBGc CA..G.....AT.....
6TBBGd T.TGTAC....AATCT
6TBGe
6BBGf T.TGT.C....AATCT
6BBGg ACACTGACTCTGCTTTTCCTTTGCAGAGCA..G.....AT.....GTGAGTCTCCCACCT...ACCAA..A.ATT.GGGGTCTTCCCATGGGATCAGC
6BBGh T.TGTAC....AATCT
6BBGi TGTGG....T.CT.GG.T.GTGGGATGT
PTBBGa GGAACAAGCAGCGCT.T
PTBBGb GGAACAAGCAGCGCT.T
PTBBGc TGT.C....A.TC..AAGATCCTAGCTTCAAA.C
15TBBGa GTTGGCGGAACAAGCAGCGCTATCGAAGCA..G.....AT.....
15TBBGb GTTT..GT.C....A.TC..A
15TBBGc GTTCCT.T.C....A.TC..A
15BBGd GTTGGCGGAACAAGCAGCGCTATCGAAGCA..G.....AT.....
15BBGe GTTT..GT.C....A.TC..A
BG8-B12 CA..G.....AT.....
BG12-B12 CA..G.....AT.....
BG9-B12 CA..G.....AT.....
BG13-B12 TT..GT.C..A.AGTC..A
BG3-B12 TT..GT.C..A.A.TC..A
BG4-B12 TT..GT.C..A.A.TC..A
BG6-B12 TT..GT.C..A.A.TC..A
BG2-B12 AG.........AT....A
BG1-B12 AC.CT..C.....AC...
BG11-B12 GT.TGTAC....A.TC..A
BG7-B12 GT.TGTAC....A.TC..A
BG5-B12 GT.TGT.C....A.TC..A
BG0-B12
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
NTBBGa CGAAGCAAAGAGATGCAATGTTGGAC
M3-B21 AGCGCAAT
M0-B21 AGCGCAAT
M4-B21 AGCGCAAT
U60-UNK AGCGCAAT
NBBGf .....TT..C.T...AAC..A..
M1-B21 AGCTTCAA AACT.A.CGA.A....T
NM-UNK AGCTTCAA AACT.A.CGA.A....T
BG10-B12 .....AT.CT..C.T...AAC...--
U98-UNK CT.GT.A..CA.AC..A..AA.....
M2-B21 AGCTTCAA AACT.ATG.A.C.AATG
NTBGb
NTBGd T TCCTTTG.AG.
NTBGe .....TT..C.T...AAC..A..
NTBGc
6TBBGa .....TT..C.T...AAC..A--
6TBBGb AGATCCTAGCT T...AAC..A--
6TBBGc ..A..CAC.T.CT..AAC...--
6TBBGd GAAGATCCTAGCT T...AAC..AT-
6TBGe G.AACG.T--
6BBGf GAAGATCCTAGCT T...AAC..AT-
6BBGg CATGGGATGATAACCTGAACCTTCTCATCGTGCGTTTCTTATTTGTTCCTTTTGCAGAGAAAACACGTT CT..AAC...--
6BBGh GAAGATCCTAGCT T...AAC..AT-
6BBGi GTTATTCCTGGCTGATTCA..TG.A...ACCT.T..CAA.C..TT
PTBBGa
PTBBGb .....G.....G
PTBBGc T..T.A..CA.ATG.A..AA.....G
15TBBGa ..A..CAC.T.CT..AAC...--
15TBBGb ..A..TT..CCT...AAC..A--
15TBBGc ..ACCCT..CCT...AAC..A--
15BBGd ..A..CAC.T.CT..AAC...--
15BBGe ..A..TT..CCT...AAC..A--
BG8-B12 ..A..CAC.T.CT..AAC...--
BG12-B12 ..A..CAC.T.CT..AAC...--
BG9-B12 ..A..CAC.T.CT..AAC...--
BG13-B12 .....TT..C.T...AAC..A--
BG3-B12 ..G..TT..CCT...AAC..A--
BG4-B12 ..ACCCT..CCT...AAC..A--
BG6-B12 ..A..TT..C.T...AAC..A--
BG2-B12 ..AA..CTGAC.....-------
BG1-B12 ..AG...G.......A.....--
BG11-B12 ..ATCCT..C.T...AAC..A--
BG7-B12 ..ATCCT..C.T...AAC..A--
BG5-B12 ..ATCCT..C.T...AAC..A--
BG0-B12 ...--
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
NTBBGa AAACACGTTCT AAAACTGGA-GGAAAAGACAGACGAAGTGGAGAATTGGA
M3-B21 .....-.....-.....
M0-B21 .....-.....-.....C....
M4-B21 .....-.....-.....
U60-UNK .....-.....-.....-.....
NBBGf G..A.T.C.-.....G....C.
M1-B21 -.....G....C.
NM-UNK -.....G....C.
BG10-B12 T.A..C.A..T..AA..T....C.TAC...
U98-UNK .T..GGAA.TC .CTAAT..A.A..C..TAT..AATGACA.....ACAAG
M2-B21 -.....AA..T......T.CA..
NTBGb .....-.....-.....
NTBGd .....-.....A.....
NTBGe G..A.T.C.-.....G....C.
NTBGc .....-.....-.....
6TBBGa AC....TG.T.....C....C.
6TBBGb T.A..C.A.TG..AA..T......T.CA..
6TBBGc A.....
6TBBGd GAAAC...-----TG..AA..T......T.CA..
6TBGe AT..G..T..G..G...C...C-------
6BBGf GAAAC...-----TG..AA..T......T.CA..
6BBGg A.....
6BBGh GAAAC...-----TG..AA..T......T.CA..
6BBGi CCTTC.AG.T.GTTTAATTCCTTCTTGGGCCCAAAGTGGTCATTGGACTCCTCCCAGAAAA..GGG.TTGG..TC.G.GTGTGA..GC..AT.GCAC...
PTBBGa .....ACT.GA.....
PTBBGb .....ACT.GA.....
PTBBGc .TT..GAA.TC .CT..T..A.A..CG.TAT..GATTAC....G.ACT.G
15TBBGa A.....
15TBBGb ACA....TG.T.....G....C.
15TBBGc ACAG..TTG.T.....G....C.
15BBGd A.....
15BBGe ACA....TG.T.....G....C.
BG8-B12 A.....
BG12-B12 A.....
BG9-B12 A.....
BG13-B12 AC.....TG.T.....C....C.
BG3-B12 ACA....TG.T..T.....G....C.
BG4-B12 AC.....TG.T.....G....C.
BG6-B12 AC.....TG.T.....G....C.
BG2-B12 .TG.G...TA.
BG1-B12 A.AC.CTAGTT..AA.TC.....GAA.A.
BG11-B12 T.A..C.A.TG..AA..T......T.CA..
BG7-B12 T.A..C.A.TG..AA..T......T.CA..
BG5-B12 T.A..C.A.TG..AA..T......T.CA..
BG0-B12 T.A....A.TT..A...T.T.....AGA..

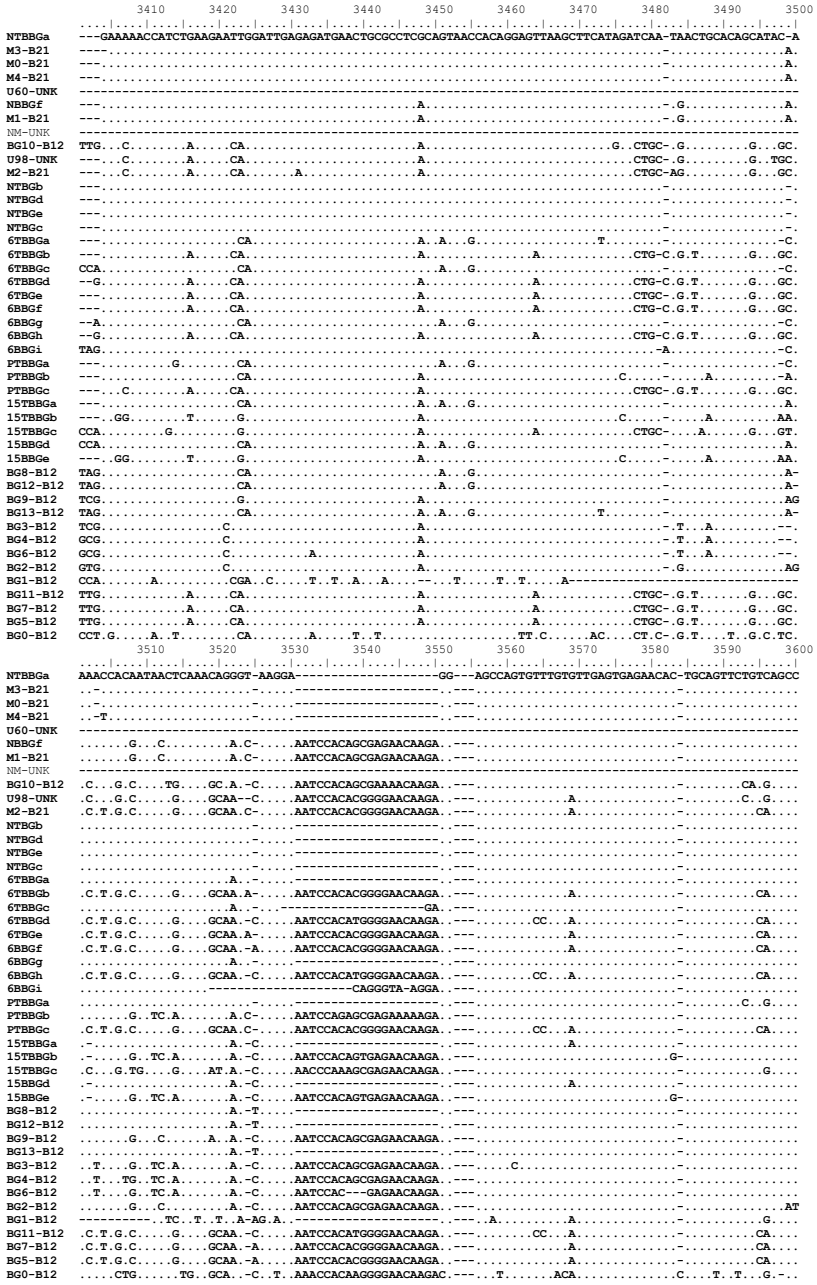

2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
NTBBGa
M3-B21
M0-B21
M4-B21
U60-UNK
NBBGf
M1-B21
NM-UNK
BG10-B12
U98-UNK
M2-B21
NTBGb
NTBGd
NTBGe
NTBGc
6TBBGa
6TBBGb
6TBBGc
6TBBGd
6TBGe
6BBGf
6BBGg
6BBGh
6BBGi
PTBBGa
PTBBGb
PTBBGc
15TBBGa
15TBBGb
15TBBGc
15BBGd
15BBGe
BG8-B12
BG12-B12
BG9-B12
BG13-B12
BG3-B12
BG4-B12
BG6-B12
BG2-B12
BG1-B12
BG11-B12
BG7-B12
BG5-B12
BG0-B12
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
NTBBGa
M3-B21
M0-B21
M4-B21
U60-UNK
NBBGf
M1-B21
NM-UNK
BG10-B12
U98-UNK
M2-B21
NTBGb
NTBGd
NTBGe
NTBGc
6TBBGa
6TBBGb
6TBBGc
6TBBGd
6TBGe
6BBGf
6BBGg
6BBGh
6BBGi
PTBBGa
PTBBGb
PTBBGc
15TBBGa
15TBBGb
15TBBGc
15BBGd
15BBGe
BG8-B12
BG12-B12
BG9-B12
BG13-B12
BG3-B12
BG4-B12
BG6-B12
BG2-B12
BG1-B12
BG11-B12
BG7-B12
BG5-B12
BG0-B12

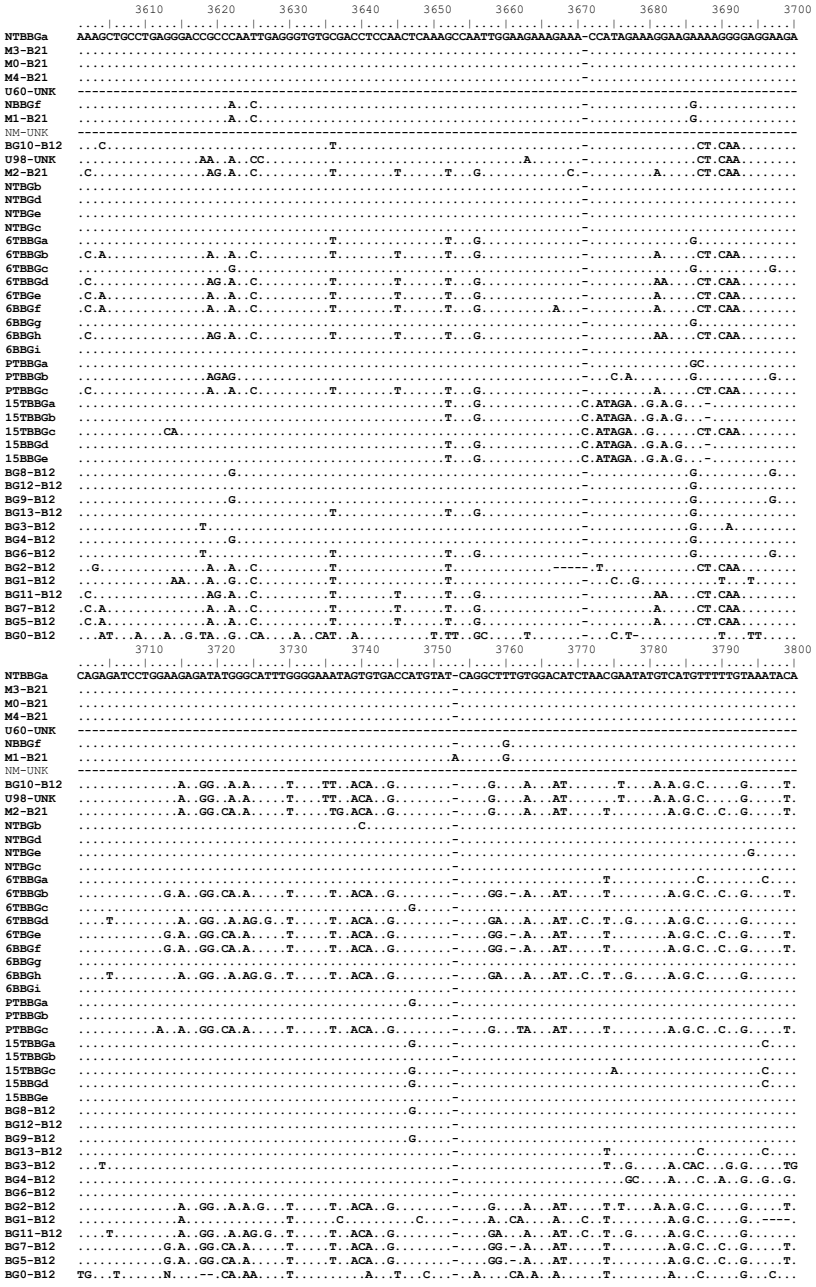

2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
NTBBGa
M3-B21
M0-B21
M4-B21
U60-UNK
NBBGf
M1-B21
NM-UNK
BG10-B12
U98-UNK
M2-B21
NTBGb
NTBGd
NTBGe
NTBGc
6TBBGa
6TBBGb
6TBBGc
6TBBGd
6TBGe
6BBGf
6BBGg
6BBGh
6BBGi
PTBBGa
PTBBGb
PTBBGc
15TBBGa
15TBBGb
15TBBGc
15BBGd
15BBGe
BG8-B12
BG12-B12
BG9-B12
BG13-B12
BG3-B12
BG4-B12
BG6-B12
BG2-B12
BG1-B12
BG11-B12
BG7-B12
BG5-B12
BG0-B12
2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
NTBBGa
M3-B21
M0-B21
M4-B21
U60-UNK
NBBGf
M1-B21
NM-UNK
BG10-B12
U98-UNK
M2-B21
NTBGb
NTBGd
NTBGe
NTBGc
6TBBGa
6TBBGb
6TBBGc
6TBBGd
6TBGe
6BBGf
6BBGg
6BBGh
6BBGi
PTBBGa
PTBBGb
PTBBGc
15TBBGa
15TBBGb
15TBBGc
15BBGd
15BBGe
BG8-B12
BG12-B12
BG9-B12
BG13-B12
BG3-B12
BG4-B12
BG6-B12
BG2-B12
BG1-B12
BG11-B12
BG7-B12
BG5-B12
BG0-B12

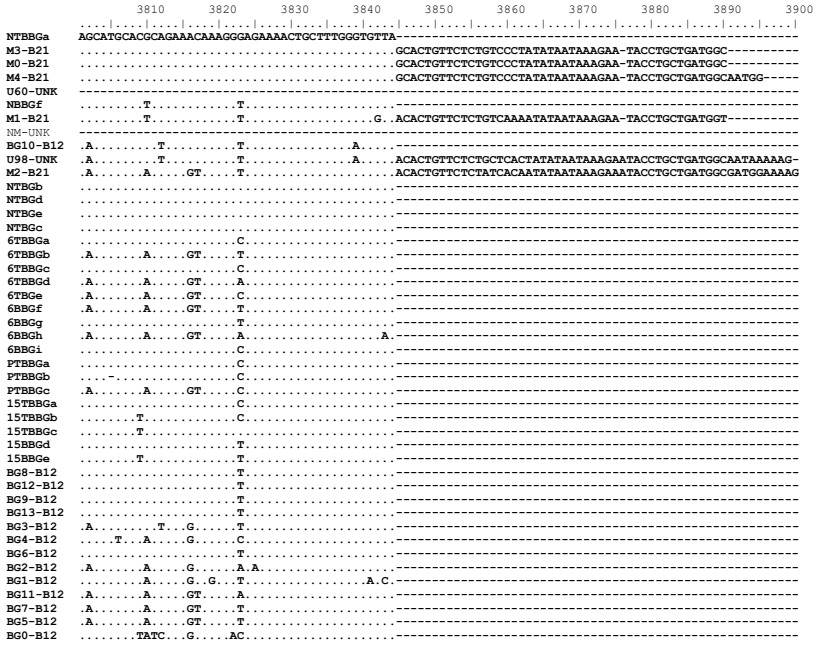

2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
NTBBGa -AACATCAAACTAAAGAATTGG-AGAAACAGCATTCACAGTTCC----------
M3-B21 -.....-.....-.....----------
M0-B21 -.....-.....-.....----------
M4-B21 -.....-.....-.....----------
U60-UNK ----------
NBBGf -C...A.....-.....GTAAGTTGCAGTCACTAAACTGAAGGAATGTGGGGTCTTCCCAAAGTCCTGCATAC
M1-B21 -C...A.....-.....----------
NM-UNK -C...A.....-.....----------
BG10-B12 -.GA.A..GT..C....C...-C.....AAC.GA.G.AG.GG----------
U98-UNK -.GA.A...GT..C....C...-C.....AAC.GA.G.AG.GG----------
M2-B21 -TGG.A...GAG...C.G...-..G...G.G.....C.GA----------
NTBGb -.....-.....-.....----------
NTBGd -.....-.....-.....----------
NTBGe -C...A.....-.....----------
NTBGc -.....-.....-.....----------
6TBBGa --...CA..G.....-.....T.....T----------
6TBBGb GTGG.A....GAG...C.G...-..G..TG.A.....C.GA----------
6TBBGc --.....-.....-.....----------
6TBBGd -GGA.-....GAG...C.G...AG...TGG.-.....--AG.----------
6TBGe G..G.CAC....T..A..G.T.-GAG...GAA...G.AGAAGTG----------
6BBGf -GGA.-....GAG...C.G...AG...TGG.-.....--AG.----------
6BBGg G.....-.....-.....----------
6BBGh -GGA.-....GAG...C.G...AG...TGG.-.....--AG.----------
6BBGi -GGA.-A.G..AG.C...G...AGA.TTGGA-....GT---G.----------
PTBBGa -.....-.....-.....----------
PTBBGb -.....-.....-.....----------
PTBBGc -.GACA...CT....A.G....-....CG.AAT.GA.G.AG.GG----------
15TBBGa -....A.G.....-.C.....----------
15TBBGb -....A.G.....-.C....A.....----------
15TBBGc -.TA.A.T.G..GCCA..C...-..CT.....----------
15BBGd -....A.G.....-.C.....----------
15BBGe -....A.G....T.....-.C....A.....----------
BG8-B12 -.....-.....-.....----------
BG12-B12 -.....-.....-.....----------
BG9-B12 -.....-.....-.....----------
BG13-B12 -G...CA..G.....-.....T.....T----------
BG3-B12 -G..GA.....-.....T----------
BG4-B12 -C...A.....-.....T----------
BG6-B12 -C...A.....-.....----------
BG2-B12 -TGG.AAG....GG.....-...G...TTTAAG.A..A.GG----------
BG1-B12 -TG..A....C....C.G...-....TT.A.....GGA----------
BG11-B12 -TGG.A....GAG...C.G...-..G..TG.G.....C.GA----------
BG7-B12 -TGG.A....GAG...C.G...-..G..TG.G.....C.GA----------
BG5-B12 -TGG.A....GAG...C.G...-..G..TG.G.....C.GA----------
BG0-B12 -TGG.A....T.G.....T.-....G..A.....A..GA----------
2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
NTBBGa ----------AGAGACACTTTCAGAATATGT-ATTTAAGTGCTGGAAAACAGA----
M3-B21 ----------.....-.....----
M0-B21 ----------.....-.....----
M4-B21 ----------.....-.....----
U60-UNK ----------
NBBGf NGGGATGAAAAAATCCCCTCTGACCATGCACTGCTTTTCTCTTTCTATTCCAG.C.....CG...G-.....T..C.....
M1-B21 ----------.C.....CG...G-.....T..C.....
NM-UNK ----------.C.....CG...G-.....T..C.....
BG10-B12 ----------.A.AT...A.G.AG.G...-GG....C.....CCT..T.T.-----
U98-UNK ----------.A.AT...A.G.AG.G...G-G....C.....CCT..T.T.-----
M2-B21 ----------G..A...G.A.G.A..GT..G-G..CGC.....AC...T.T.-----
NTBGb ----------.....-.....----
NTBGd ----------.....-.....----
NTBGe ----------.C.....CG...G-.....T..C.....
NTBGc ----------.....-.....----
6TBBGa ----------.C.....C.....-...C.....----
6TBBGb ----------G..A...G.ACG.A..GT..G-G..CGC.....C...T.T.-----
6TBBGc ----------.....-.....----
6TBBGd ----------TGAG..A...G.ACG.A..GT...-GG..CGC.....C...T.T..AGA-
6TBGe ----------GA.A....C..A.A..G..TG-G.A..C.....CCT..T.T.-----
6BBGf ----------TGAG..A...G.A.G.A..GT...-GG..CGT.....C...T.T..AGA-
6BBGg ----------.....-.....----
6BBGh ----------TGAG..A...G.ACG.A..GT...-GG..CGC.....C...T.T..AGA-
6BBGi ----------TGA.A.A.G..AG.G.AG.G....GG..ATG.CTT...G.T.T..GTAA
PTBBGa ----------.....-.....----
PTBBGb ----------.....-.....T..T.....----
PTBBGc ----------...A....C..A.A..G.CTG-G.AG.C.....CCT..T.T.-----
15TBBGa ----------.C.....-T.....----
15TBBGb ----------.....-G....C.....----
15TBBGc ----------.....-G.....----
15BBGd ----------.C.....-T.....----
15BBGe ----------.....-G....C.....----
BG8-B12 ----------.....-T.....----
BG12-B12 ----------.....-T.....----
BG9-B12 ----------.....-T.....----
BG13-B12 ----------.C.....C.....-T...C.....----
BG3-B12 ----------.C.....C...C.....-G...C...A....T.....----
BG4-B12 ----------.C.....C...C.....-G.....T.....----
BG6-B12 ----------.C.....CG....-G.....T...C.....----
BG2-B12 ----------..CA....A..GTAG.AGG.-G..A.....A...T.T.-----
BG1-B12 ----------GT.AG.----------..C.G.C----------TT.A.----
BG11-B12 ----------G..A...G.ACG.A..GT...-GG..CGC.....C...T.T.-----
BG7-B12 ----------G..A...G.ACG.A..GT...-GG..CGC.....C...T.T.-----
BG5-B12 ----------G..A...G.A.G.A..GT...-GG..CGT.....C...T.T.-----
BG0-B12 ----------..GA...T.A.AT...G...-GT.....C.G.T.T.-----
2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
NTBBGa ----------
M3-B21 ----------
M0-B21 ----------
M4-B21 ----------
U60-UNK ----------
NBBGf ----------
M1-B21 ----------
NM-UNK ----------
BG10-B12 ----------
U98-UNK ----------
M2-B21 ----------
NTBGb ----------
NTBGd ----------
NTBGe ----------
NTBGc ----------
6TBBGa ----------
6TBBGb ----------
6TBBGc ----------
6TBBGd --------CACAACTTAAAAA--GTT----------
6TBGe ----------
6BBGf --------CACAACTTAAAAA--GTT----------
6BBGg ----------
6BBGh --------CACAACTTAAAAA--GTT----------
6BBGi GTCTCCCTCCCAACATGGAAGGAATTTATGGTCTTAACATGGGATCAGCCATGGGATGATCATCTGACCCCTCTCATCATGCAATTCATATTTGTTCCTT
PTBBGa ----------
PTBBGb ----------
PTBBGc ----------
15TBBGa ----------
15TBBGb ----------
15TBBGc ----------
15BBGd ----------
15BBGe ----------
BG8-B12 ----------
BG12-B12 ----------
BG9-B12 ----------
BG13-B12 ----------
BG3-B12 ----------
BG4-B12 ----------
BG6-B12 ----------
BG2-B12 ----------
BG1-B12 ----------
BG11-B12 ----------
BG7-B12 ----------
BG5-B12 ----------
BG0-B12 ----------
2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
NTBBGa ------AGAAAATGGTTACAAAACTGGAGGAACACTGTGAATGGATGGTGAGAAGG----------
M3-B21 ------.....----------
M0-B21 ------.....----------
M4-B21 ------.....----------
U60-UNK ----------
NBBGf ------.....C.....T.....A.....----------
M1-B21 ------.....C.....T.....A.....----------
NM-UNK ------.....C.....T.....A.....----------
BG10-B12 ------.....ATA.AGT.G.....A...AAC...GCACG...A..ATC..----------
U98-UNK ------.....ATA.AGT.G.....A...AAC...GCACG...A..ATC..AAATCAGAGCTGAAGA-----AACAGTATGAAAATTTGGCTTC
M2-B21 -------..C.CAAC...A...GT....A.CG.AAT...GAAG...A..A.CAC----------
NTBGb ------.....----------
NTBGd ------.....----------
NTBGe ------.....C.....T.....A.....----------
NTBGc ------.....----------
6TBBGa ------..G..C.....T.....CA.....-AGAG.A.G----------
6TBBGb ------...C.CAAC...A...GT....A.CG.AAT...GAAG...-AGA..CACCTTAAAAAGATTGGTATACGTGCTC-CTAATCTGAG---GCT
6TBBGc ------.....-TGAG.A.GAATGTAAAGTTGGAGGCAGCAGCTGTAAAAGTGGTGAGA----
6TBBGd ----GG....CGAAA..GA.G..G....A....CT.A..AA...T.GT.-T.C.TGCTCCTAATCTGAGGCTACACAT-GGCAGAACTGGTGGAT---
6TBGe ------G.CT.CACA.GG..G.....T...T..T..AGT...GCAG...-AGA.TC.G----------
6BBGf ----GG....CGAAA..GA.G..G....A....CT.A..AA...T.GT.-T.C.TGCTCCTAATCTGAGGCTACACAT-GGCAGAACTGGTGGAT---
6BBGg ------.....-TGAG.A.GAATGTAAAGTTGGAGGCAGCAGCTG-TAAAAGTGGT----GAG
6BBGh ----GG....CGAAA..GA.G..G....A....CT.A..AA...T.GT.-T.C.TGCTCCTAATCTGAGGCTACACAT-GGCAGAACTGGTGGAT---
6BBGi TTGCAG.....C...C.G..G.....A.....C....GA.....G..-C.A.GGATTTAAAGTTGGAGCGACTAGCTGCCA-AACTGGAACAT---
PTBBGa ------.....AATGTAAAGTTGG--AGATACCAGCTGTAAAAGTGG----TGAG
PTBBGb ------.....C.....CA.....AATGCAAGGTTGG--AGGCAGCAGCTGTAAAAG----------
PTBBGc ------..C..CACA.GG..G.....T...T..AGC....GCAG...A..ATC..AAATCAGAGCTGAGTAAGTCGCAGTCACTGAACTGA--------
15TBBGa ------.....----------
15TBBGb ------..G..C.....CA.....A.....----------
15TBBGc ------..G..C.....T.....C.....C.....----------
15BBGd ------.....----------
15BBGe ------..G..C.....CA.....A.....----------
BG8-B12 ------.....----------
BG12-B12 ------.....----------
BG9-B12 ------.....----------
BG13-B12 ------..G..C.....T.....CA.....A.....----------
BG3-B12 ------.....C.....C....CA.....A.....----------
BG4-B12 ------.....C.....C....CA.....A.....----------
BG6-B12 ------.....C.....T.....A.....----------
BG2-B12 ------..G..CAA.C.....T...T.....AAC.A..GAATCA..AGT.G----------
BG1-B12 ------.CTG.CCAAG.CTG.C..CATCCA..ATAAC.TC..A.GCTA...A..AT----------
BG11-B12 ------...C.CAAC...A...GT....A.CG.AAT...GAAG...A..A.CAC----------
BG7-B12 ------...C.CAAC...A...GT....A.CG.AAT...GAAG...A..A.CAC----------
BG5-B12 ------...C.CAAC...A...GT....A.CG.AAT...GAAG...A..A.CAC----------
BG0-B12 ------.....CAA...G..G.....----------


2810 2820 2830 2840 2850 2860 2870 2880 2890 2900
NTBBGa ----AATGTAAAGTTGGAGATACCAGCTGTAAAAGTGG----------GGCAACAAGCTAAAGAATCA
M3-B21 ----.....GTAAAAGTGG----------.....
M0-B21 ----.....GTAAAAGTGG----------.....
M4-B21 ----.....GTAAAAGTGG----------.....
U60-UNK ----------
NBBGf ----.....T.C...----------.....
M1-B21 ----.....GTAATACTGG----------.....
NM-UNK ----.....GTAATACTGG----------.....
BG10-B12 ----..ATC.G..C..A...A..AGTA..A...TT...CTTCACATGCTTCAGAGCTGAAGAAACAAGCTGAAGTACTGGA.G.....G..C..CTG
U98-UNK ACATGC.TC.G..C..A...A..A....C.GT.C..A----------A.G.G.....G..C..CTG
M2-B21 ----CT.AA....A.T.GT...GT...CCT..TC..AAGCTACACATGGCAGAACTGG----------T.G.T.....G...C.GTG
NTBGb ----.....----------.....
NTBGd ----.....----------.....
NTBGe ----.....T.C...----------.....
NTBGc ----.....----------.....
6TBBGa ----....C..........GC.G..........C..----------.A..CA.....
6TBBGb ACAC.TG.C.G.AC...T.G----------.T...T.G...C.GTG
6TBBGc -AGG.....GC.G..----------.A..CA.....
6TBBGd ----------C-...T.G...C.GTG
6TBGe ----..ATC.G...G.A...G.TATTTGAC...TA.A.----------.TTT..GT...G--------
6BBGf ----------C-...T.G...C.GTG
6BBGg AAGG.....GC.G..----------.A..CA.....
6BBGh ----------C-...T.G...C.GTG
6BBGi ----------C-..A.....TG
PTBBGa AAGG.....----------.....
PTBBGb ----------
PTBBGc -------.CG..T....G.TCTT.CCAA.GG.C..CA----------T.
15TBBGa ----.....T.....GC.G.....TGAGAAGGAATGTAAAGTTGGAGGCAGCAGCTGTAAAAGTGG.A..CA.....
15TBBGb ----.....GC.....----------
15TBBGc ----.....----------
15BBGd ----.....T.....GC.G.....TGAGAAGGAATGTAAAGTTGGAGGCAGCAGCTGTAAAAGTGG.A..CA.....
15BBGe ----.....GC.....----------
BG8-B12 ----.....GC.G.....TGAGAAGGAATGTAAAGTTGGAGGCAGCAGCTGTAAAAGTGG.A..CA.....
BG12-B12 ----.....GC.G.....----------.A..CA.....
BG9-B12 ----.....GC.G.....----------.A..CA.....
BG13-B12 ----....C.....GC.G.....C..----------.A..CA.....
BG3-B12 ----.....GC.G.....----------.....
BG4-B12 ----.....GC.G.....----------.....
BG6-B12 ----.....T.....T.C...----------.....
BG2-B12 ----------..T..G.....
BG1-B12 ----CCCCCC.G.CC.TGA.CTA.TCTCC.CTTTC.AACCC----------AG.GA..CACC.C...G..
BG11-B12 ----CT.AA....A.T.GT...GT...CCT..TC..AGGCTACACATGGCAGAACTGGT----------.G.T...T.G...C.GTG
BG7-B12 ----CT.AA....A.T.GT....GT...CCT..TC..AGGCTACACATGGCAGAACTGGT----------.G.T...T.G...C.GTG
BG5-B12 ----CT.AA....A.T.GT....GT...CCT..TC..AGGCTACACATGGCAGAACTGGT----------.G.T...T.G...C.GTG
BG0-B12 ----------A....GTGAGT.TT.C----------C...A.C...G.AT.
2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
NTBBGa GAGGAACAGAAA----------TCGGAGCTGA----------
M3-B21 .....TCGGAGCTGA----------
M0-B21 .....TCGGAGCTGA----------
M4-B21 .....TCGGAGCTGA----------
U60-UNK ----------
NBBGf ...----------
M1-B21 ...----------
NM-UNK ...----------
BG10-B12 ...ATT.....TT----------
U98-UNK ...ATT.....TTCACTGTCGAAGATACGCAAT----------
M2-B21 ...A...G....TCAGAGCTGAAGAGCT----------
NTBGb .....----------.....
NTBGd .....----------.....GTAAGTTGCAGTCAC
NTBGe ...----------
NTBGc .....----------.....----------
6TBBGa ...A.....TCGGAGCTGA----------
6TBBGb ...A.T.G...----------.A...TG.----------
6TBBGc ...A.....TCGGAGCTGAAGGAGCGCCATGAGGAGATGGGGCAACAAGCTAAAGAATCAGAGAAACAGAAA.....----------
6TBBGd ...A.T.G...----------.A...TG..GTAAGTCGAAGTCAC
6TBGe ----------CAGAGCTGA----------
6BBGf ...A.T.G...----------.A...TG.----------
6BBGg ...A.....----------.....----------
6BBGh ...A.T.G...----------.A...TG..GTAAGTCGAAGTCAC
6BBGi ...A.....C.T----------.AC..T.CCGTAAGTTGCAGTCAC
PTBBGa ...A.....TCGGAGCTGAAGGAGCGCCATGAGGAGATGGGGCAACAAGCTAAAGA---A----------
PTBBGb ----------TGGGACACAAAGCTAAAGA---A----------
PTBBGc CG...TG.A...TCCCCTCTGACCATGCGCT----GATTTTGTCTTTCTATTCCAGAGAGATA----------
15TBBGa ...A.....TCGGAGCTGA----------
15TBBGb ----------
15TBBGc ----------
15BBGd ...A.....TCGGAGCTGA----------
15BBGe ----------
BG8-B12 ...A.....T----------
BG12-B12 ...A.....T----------
BG9-B12 ...A.....T----------
BG13-B12 ...A.....T----------
BG3-B12 ...A.....T----------
BG4-B12 ...A.G....T----------
BG6-B12 ...A.....T----------
BG2-B12 .....CTGCG----------
BG1-B12 A..AGG.GATGGT----------
BG11-B12 ...A.T.G....T----------
BG7-B12 ...A.T.G....T----------
BG5-B12 ...A.T.G....T----------
BG0-B12 TG...TTTCCC.C----------
3010 3020 3030 3040 3050 3060 3070 3080 3090 3100
NTBBGa ----------AGGAGCAC
M3-B21 ----------AGGAGCACCAT----------
M0-B21 ----------AGGAGCACCAT----------
M4-B21 ----------AGGAGCACCAT----------
U60-UNK ----------
NBBGf ----------
M1-B21 ----------
NM-UNK ----------
BG10-B12 ----------
U98-UNK ----------
M2-B21 ----------
NTBGb ----------.....
NTBGd TGAAATGAGGGAATGTGGGGTCTTCCCAAAGTCCTGCGTATGGGATGAAAAATCCCCTCTGACCATGCACTGCTTTTCTCCTCCTTTGCCAG.....
NTBGe ----------
NTBGc ----------.....
6TBBGa ----------.....G.
6TBBGb ----------..AGAT.T
6TBBGc ----------.....G.
6TBBGd TGAACTGAGCGAATTTGGGGTCTTCCCAAGGGACAGCATACGGGATGAAAAATCCCCTCTGATCATGCACTGCTTTTGTCTTTCTATTCCAG..AGAT.T
6TBGe ----------.AA.AA..
6BBGf ----------..AGAT.T
6BBGg ----------.....G.
6BBGh TGAACTGAGCGAATTTGGGGTCTTCCCAAGGGACAGCATACGGGATGAAAAATCCCCTCTGATCATGCACTGCTTTTGTCTTTCTATTCCAG..AGAT.T
6BBGi TAAACTGAAGGAATGTGGGGTCTTCCCAAAGTCCTGCATATGGGATGAAAAATCCCCTCTGACCATGCACTGCTTTTCTCTTTCTATTCCAG..AGA...
PTBBGa ----------TCAGAGAAACAGAAAT----------CGGAGCTGA.....G.
PTBBGb ----------TCAGAGAAACAGAAAT----------CGGAGCTGA.....G.
PTBBGc ----------TTTGACAAATATAGGTTTACGTGCTGCAGAGCTGA.AA.AT..
15TBBGa ----------.....G.
15TBBGb ----------
15TBBGc ----------
15BBGd ----------.....G.
15BBGe ----------
BG8-B12 ----------
BG12-B12 ----------
BG9-B12 ----------
BG13-B12 ----------
BG3-B12 ----------
BG4-B12 ----------
BG6-B12 ----------
BG2-B12 ----------
BG1-B12 ----------
BG11-B12 ----------
BG7-B12 ----------
BG5-B12 ----------
BG0-B12 ----------
3110 3120 3130 3140 3150 3160 3170 3180 3190 3200
NTBBGa CATGAGGAGACGGGGCAA----CAAGCTAAAGAATCAGAGAAACAGAAAT---------CGGAGCTGAAGGAGCGCCATGAGGAGA----------
M3-B21 ---------..G.A.A.GGGGCAA.....----------.....----------
M0-B21 ---------..G.A.A.GGGGCAA.....----------.....----------
M4-B21 ---------..G.A.A.GGGGCAA.....----------.....----------
U60-UNK ----------.....----------.....----------
NBBGf ----------.....----------.....----------
M1-B21 ----------.....----------.....----------
NM-UNK ----------.....----------.....----------
BG10-B12 ----------ACT.T....ATA...A..----------
U98-UNK --------A.ACATAGGGAGAGAA.GAA.G..ATG.TG.....A.CTG---------TA..A.-------.A.AAAC...A..AT----------
M2-B21 ----------..TT..C...T.T.GG.TTACGTGCTG.A.....AA.ATA.AT..CA.CAC----------
NTBGb .....----------.....----------.....----------
NTBGd .....----------.....----------.....----------
NTBGe ----------.....----------.....----------
NTBGc .....----------.....----------.....----------
6TBBGa .....T..----------.....----------
6TBBGb TTGACAA.T.TA..TTT.------.GT...G----------.A.....AA.A----------AACGTTGCAGAA
6TBBGc .....T..----------.....----------
6TBBGd TTGACAA.T.T-A..TTTACGTG-----.GC...GCTGA.A...A.CGTTG---------.A..A....GT...TCT.T----------G
6TBGe GT..CA..ACT.A----------.....----------
6BBGf TTGACAA.T.T-A..TTTACGTG-----.GC...GCTGA.A...A.CGTTG---------.A..A.....A.A----------
6BBGg .....T.....----------.....----------CGCCATGAGGAG
6BBGh TTGACAA.T.T-A..TTTACGTG-----.GC...GCTGA.A...A.CGTTG---------.A..A....GT...TCT.T----------G
6BBGi TT.C..A.T.T-.TATTTAAGTG-----.GG.A..CAGA....AT.GTTA---------.AA.A...G....A.A.TG...ATG..TGGTGAGAAGGA
PTBBGa .....T..----------.....----------
PTBBGb .....T..----------.....----------
PTBBGc GT..CA..ACT.A----------.....----------
15TBBGa .....T.....----------.....---------A.....----------
15TBBGb ----------.....----------.....---------.C.....----------
15TBBGc ----------.....----------.....---------.G.....----------
15BBGd .....T.....----------.....---------.....A.....----------
15BBGe ----------.....----------.....---------.C.....----------
BG8-B12 ----------.....----------
BG12-B12 ----------.....----------
BG9-B12 ----------.....----------
BG13-B12 ----------.....----------
BG3-B12 ----------.....----------
BG4-B12 ----------.....----------
BG6-B12 ----------.....----------
BG2-B12 ----------AACG.A..G.ATTAA.T.T.----------
BG1-B12 ----------ATATAAA.TCT..TTAT.C.----------
BG11-B12 ----------.A...TG....AGATATTTG----------
BG7-B12 ----------.A...TG....AGATATTTG----------
BG5-B12 ----------.A...TG....AGATATTTG----------
BG0-B12 ----------G...TGAC...CT.TC..TC----------


3210 3220 3230 3240 3250 3260 3270 3280 3290 3300
NTBBGa
M3-B21
M0-B21
M4-B21
U60-UNK
NBBGf
M1-B21
NM-UNK
BG10-B12
U98-UNK
M2-B21
NTBGb
NTBGd
NTBGe
NTBGc
6TBBGa
6TBBGb
6TBBGc
6TBBGd
6TBGe
6BBGf
6BBGg
6BBGh
6BBGi
PTBBGa
PTBBGb
PTBBGc
15TBBGa
15TBBGb
15TBBGc
15BBGd
15BBGe
BG8-B12
BG12-B12
BG9-B12
BG13-B12
BG3-B12
BG4-B12
BG6-B12
BG2-B12
BG1-B12
BG11-B12
BG7-B12
BG5-B12
BG0-B12
3310 3320 3330 3340 3350 3360 3370 3380 3390 3400
NTBBGa
M3-B21
M0-B21
M4-B21
U60-UNK
NBBGf
M1-B21
NM-UNK
BG10-B12
U98-UNK
M2-B21
NTBGb
NTBGd
NTBGe
NTBGc
6TBBGa
6TBBGb
6TBBGc
6TBBGd
6TBGe
6BBGf
6BBGg
6BBGh
6BBGi
PTBBGa
PTBBGb
PTBBGc
15TBBGa
15TBBGb
15TBBGc
15BBGd
15BBGe
BG8-B12
BG12-B12
BG9-B12
BG13-B12
BG3-B12
BG4-B12
BG6-B12
BG2-B12
BG1-B12
BG11-B12
BG7-B12
BG5-B12
BG0-B12


3410 3420 3430 3440 3450 3460 3470 3480 3490 3500
NTBBGa
M3-B21
M0-B21
M4-B21
U60-UNK
NBBGf
M1-B21
NM-UNK
BG10-B12
U98-UNK
M2-B21
NTBGb
NTBGd
NTBGe
NTBGc
6TBBGa
6TBBGb
6TBBGc
6TBBGd
6TBGe
6BBGf
6BBGg
6BBGh
6BBGi
PTBBGa
PTBBGb
PTBBGc
15TBBGa
15TBBGb
15TBBGc
15BBGd
15BBGe
BG8-B12
BG12-B12
BG9-B12
BG13-B12
BG3-B12
BG4-B12
BG6-B12
BG2-B12
BG1-B12
BG11-B12
BG7-B12
BG5-B12
BG0-B12
3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
NTBBGa
M3-B21
M0-B21
M4-B21
U60-UNK
NBBGf
M1-B21
NM-UNK
BG10-B12
U98-UNK
M2-B21
NTBGb
NTBGd
NTBGe
NTBGc
6TBBGa
6TBBGb
6TBBGc
6TBBGd
6TBGe
6BBGf
6BBGg
6BBGh
6BBGi
PTBBGa
PTBBGb
PTBBGc
15TBBGa
15TBBGb
15TBBGc
15BBGd
15BBGe
BG8-B12
BG12-B12
BG9-B12
BG13-B12
BG3-B12
BG4-B12
BG6-B12
BG2-B12
BG1-B12
BG11-B12
BG7-B12
BG5-B12
BG0-B12

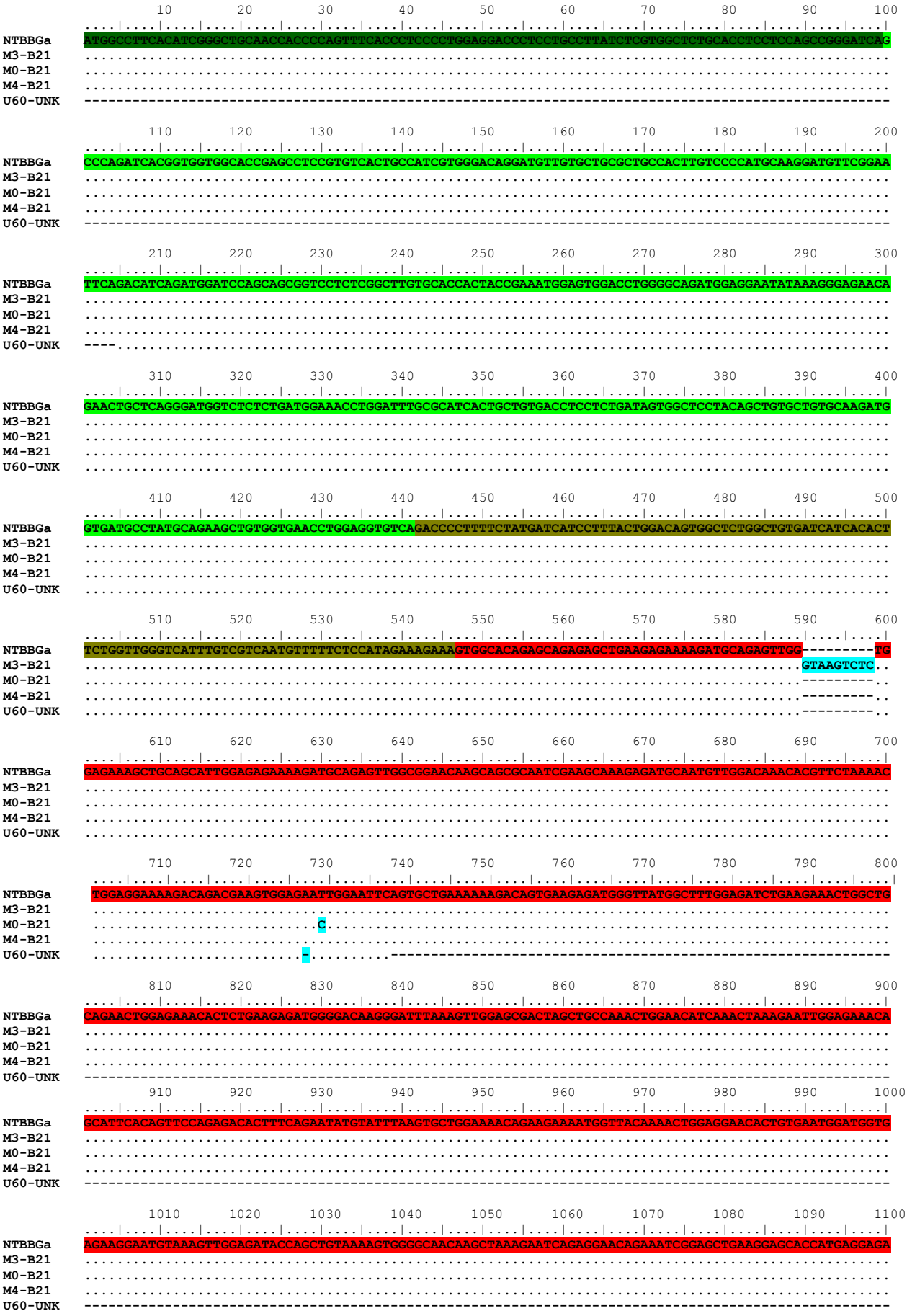


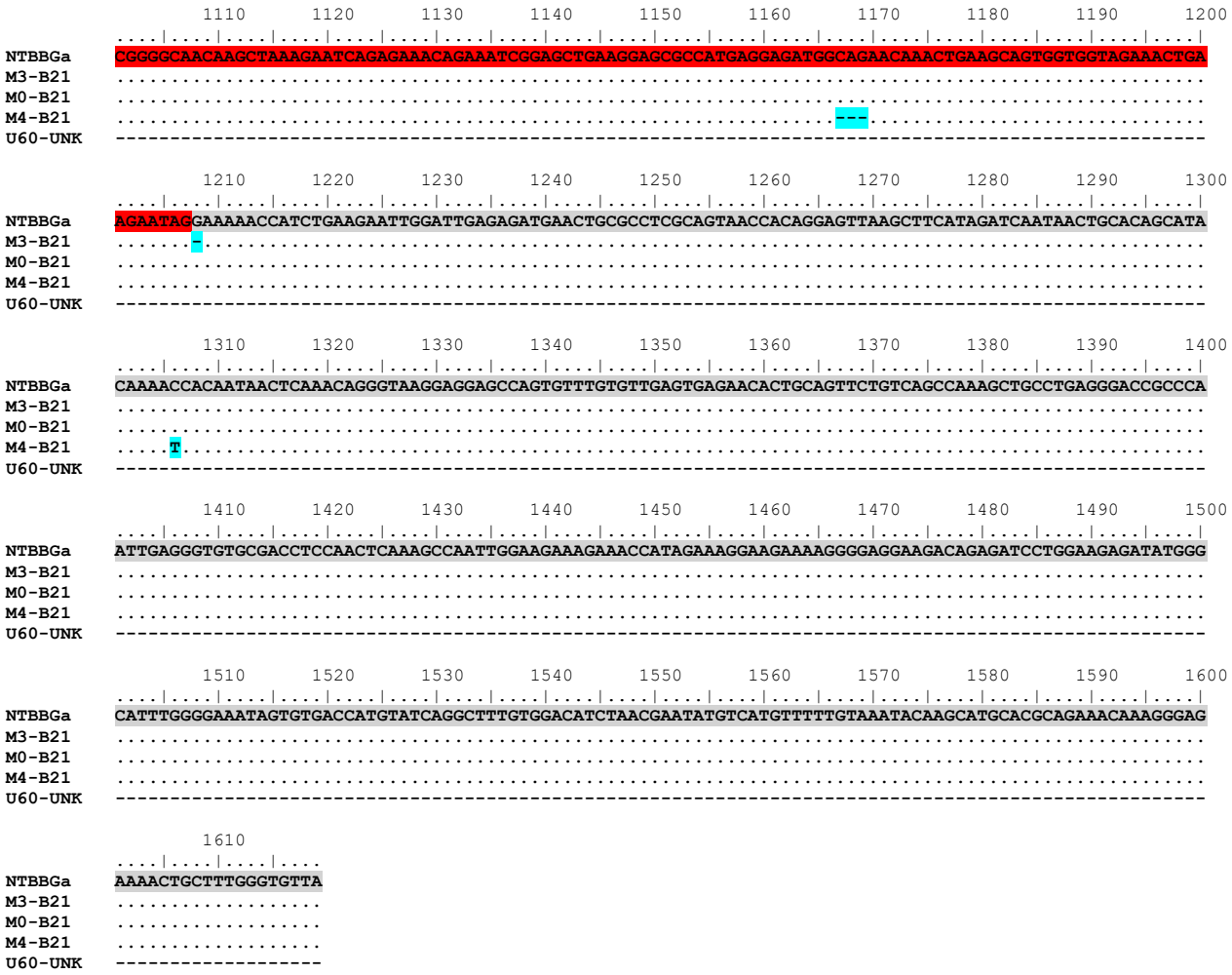


Appendix N. Alignment of cDNA sequences of all BG genes mentioned in this project. These cDNA sequences include the ones downloaded from GenBank, the full-length conceptual transcripts for all 23 genes found in T and B cells from four chicken lines, and all 14 BG genes of the B12 haplotype from Salomonsen *et al*., 2014. Names of transcripts downloaded from GenBank follow the convention: short representative names listed on table 5.1, and the B haplotype. Names of the transcripts follow the convention: “N” for line N, “T” for T cells, “B” for B cells, “BG” and the letter representing the exon 2 sequence. Names of the gene follow the convention: “BG” and the number of the gene locus for the B12 haplotype.

This page intentionally left blank

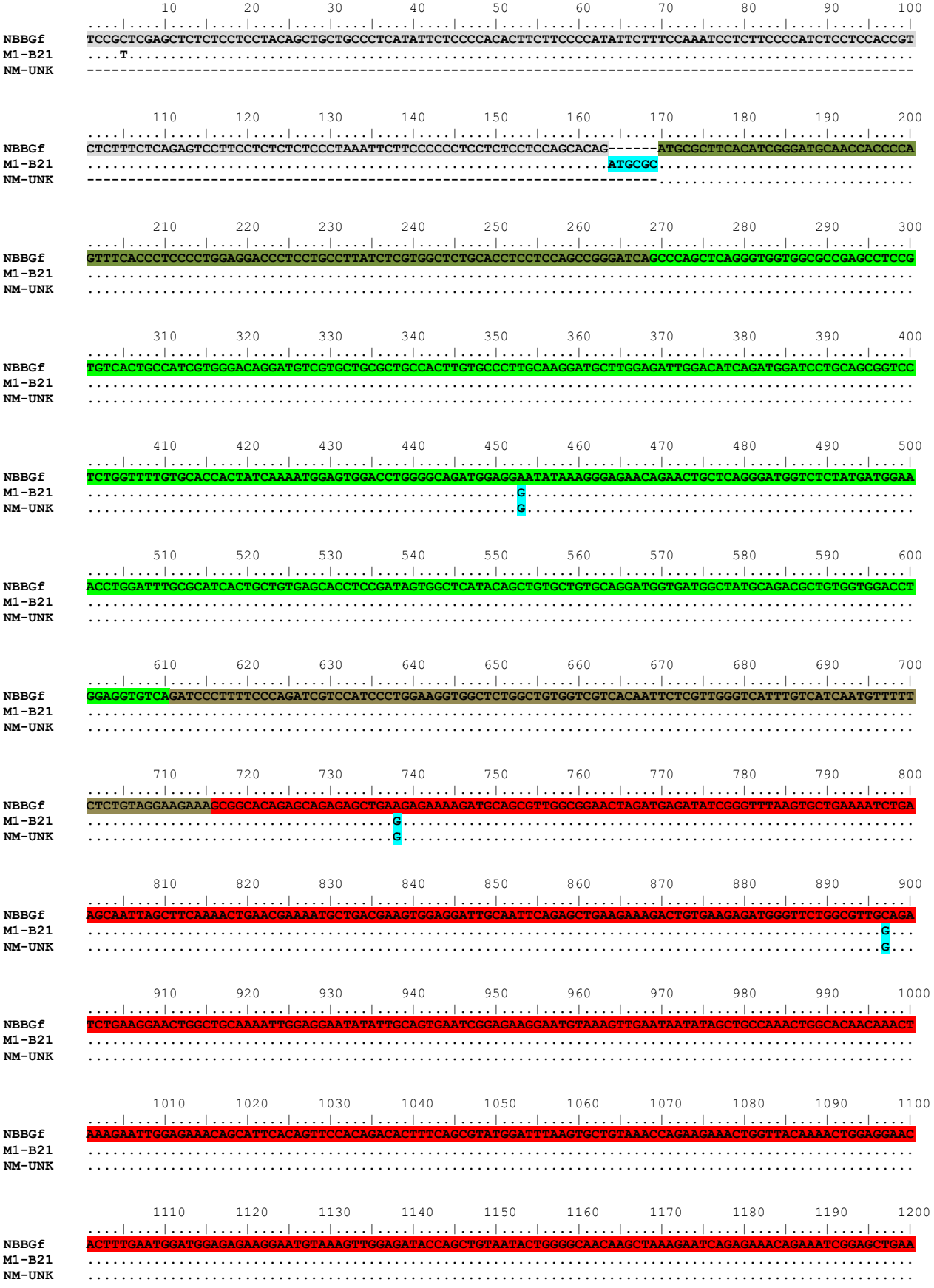
Appendix O.

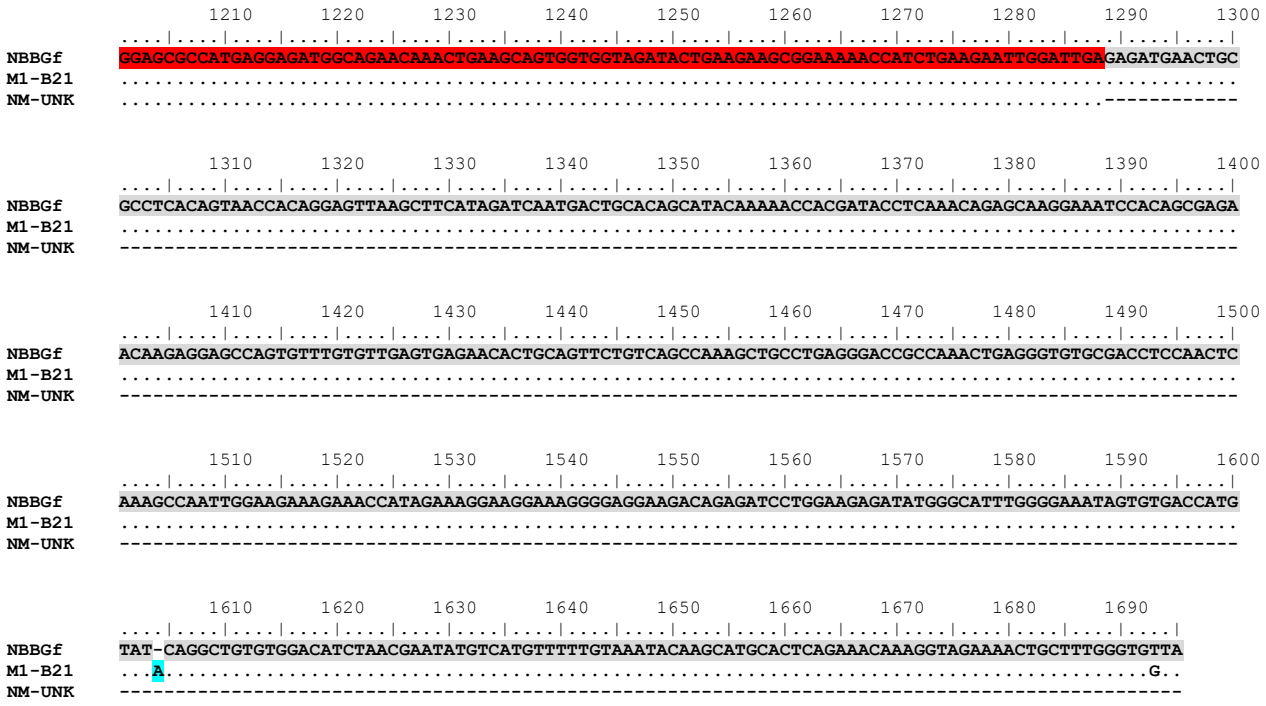




Appendix O. Nucleotide sequence alignment showing four BG transcripts downloaded from GenBank with three from B21 haplotype and one from unknown haplotype are the same gene as NTBBGa identified in this project. The top sequence is the conceptual cDNA sequence (that is exons without introns) of NTBBGa gene found in line N (B21). The rest of the four BG transcripts are downloaded from GenBank with their background information summarized in table 5.1. The comparison clearly shows that these five transcripts above are from the same BG gene with a few differences labeled blue in cytoplasmic tail region and 3' UTRs, which are most likely due to the alternatively usage of splice site (M4-B21) and PCR artifacts. Colours indicate different coding regions (grey, 5'UTR and 3'UTR; dark green, signal sequence; light green, Ig-V domain; brown, transmembrane region; red, cytoplasmic tail region).

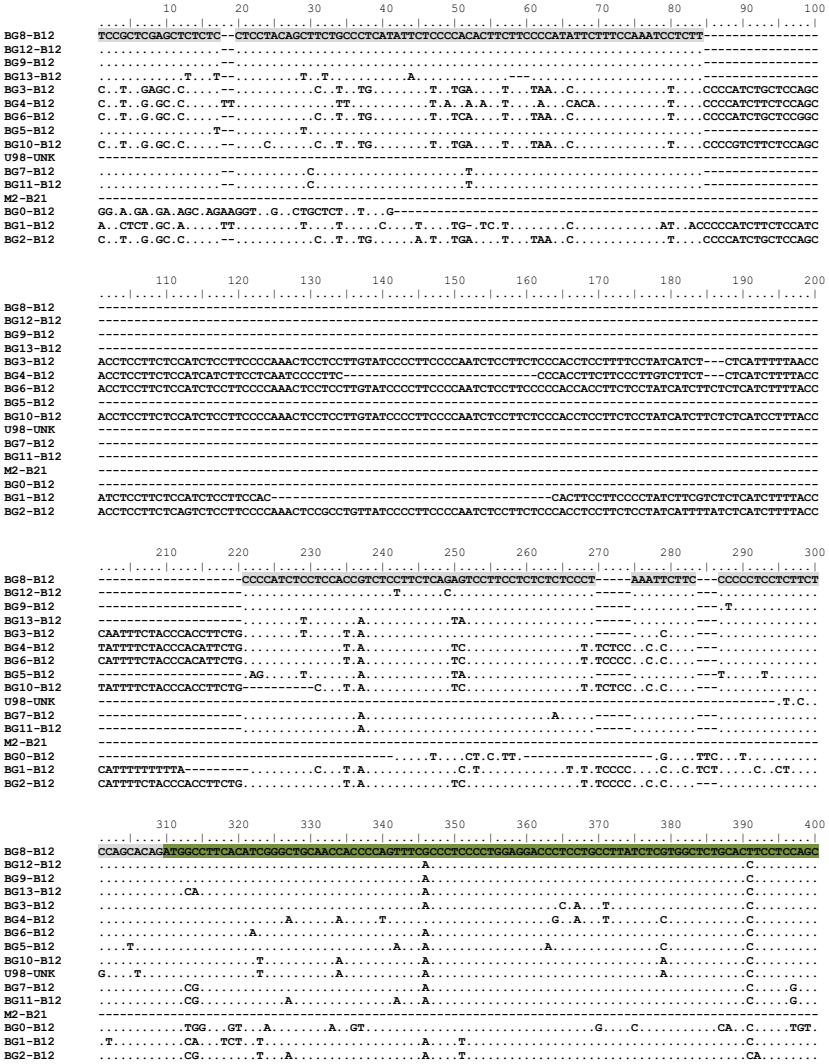
Appendix P.

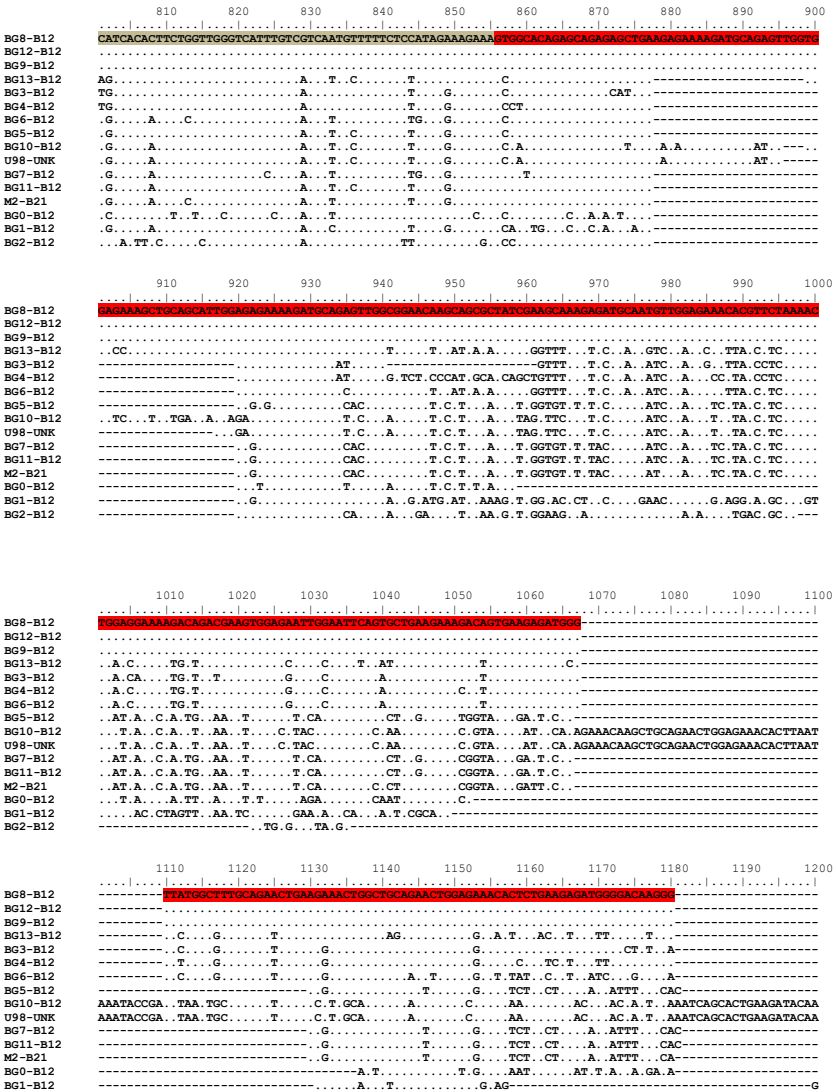


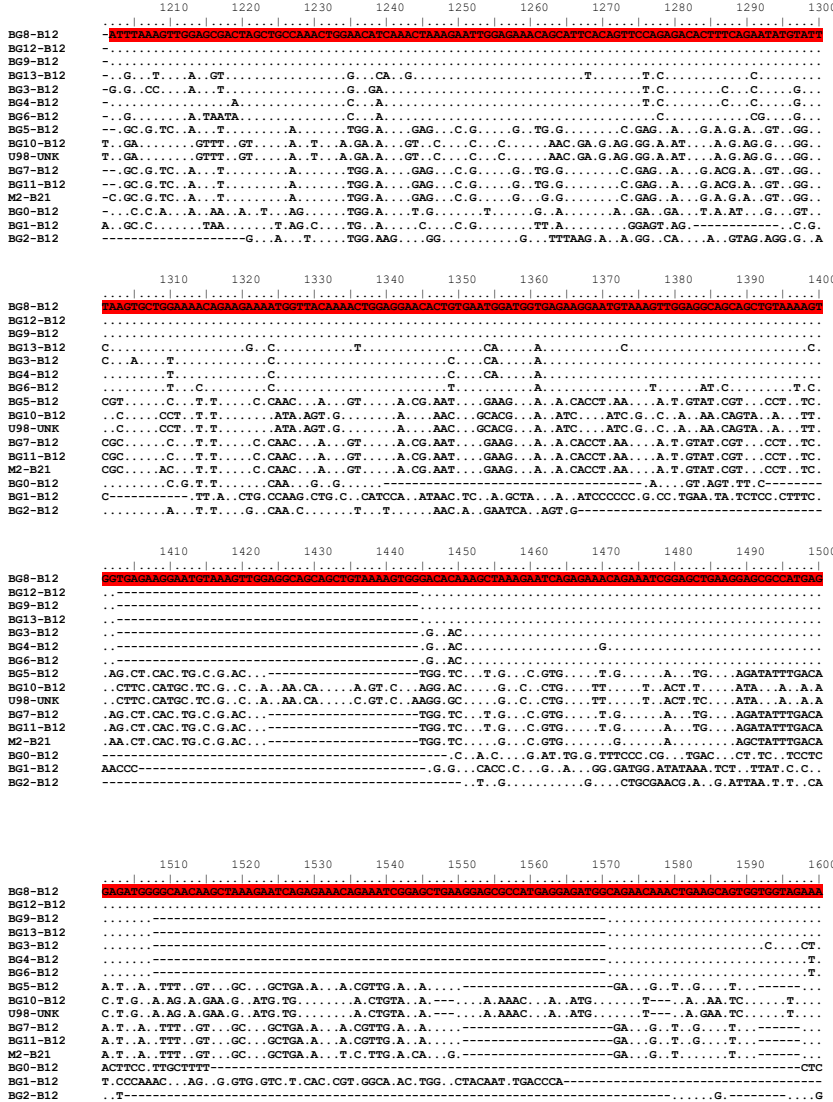


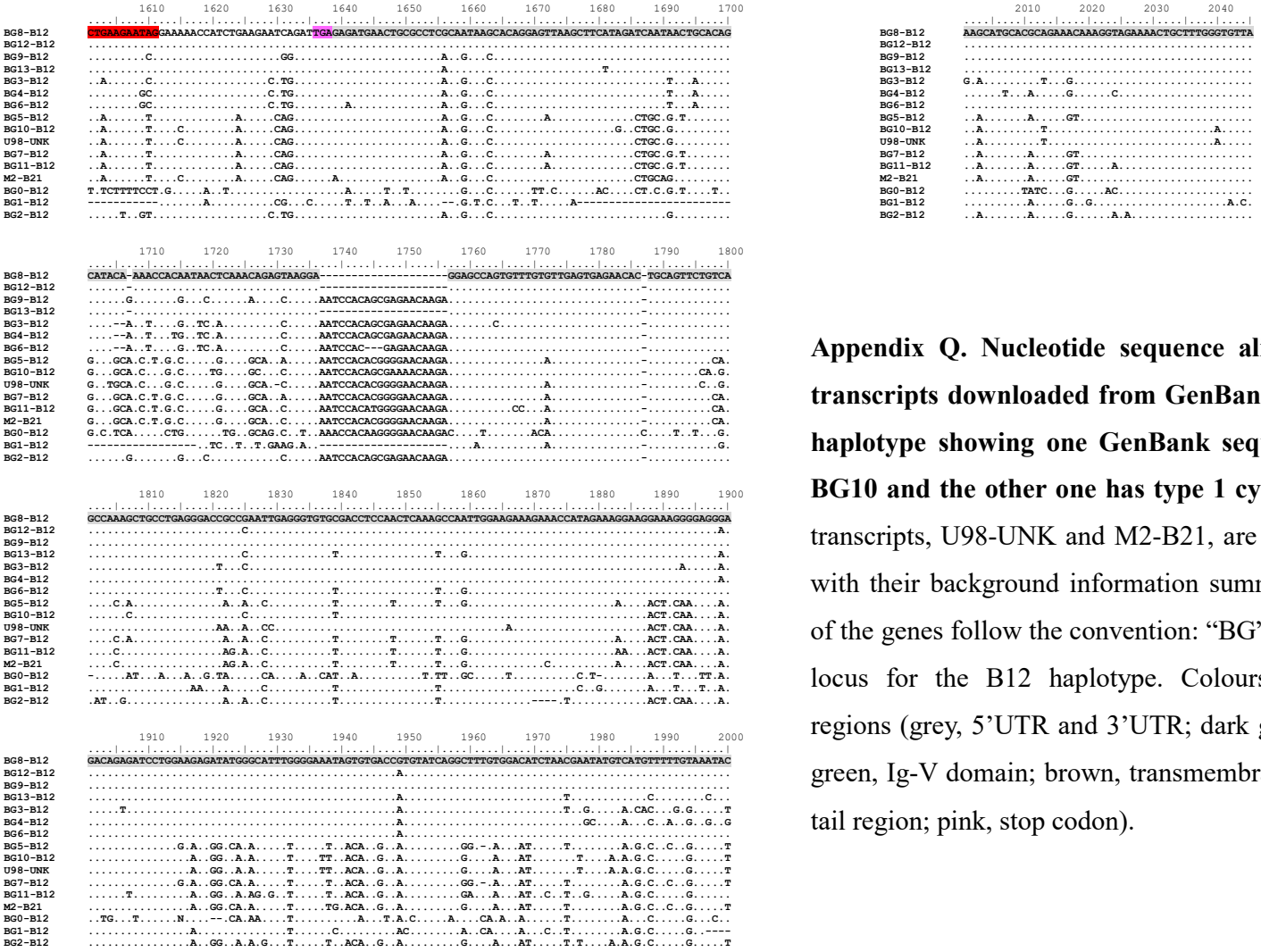
Appendix P. Nucleotide sequence alignment showing two BG transcripts downloaded from GenBank with one from B21 haplotype and one from unknown haplotype are the same gene as NBBGf identified in this project. The top sequence is the conceptual cDNA sequence (that is exons without introns) of NBBGf gene found in line N (B21). The rest of the two BG transcripts are downloaded from GenBank with their background information summarized in table 5.1. The comparison shows that these three transcripts above are most likely from the same BG gene with a few nucleotide differences labeled blue, which are most likely due to the PCR artifacts either in GenBank sequences or our sequence. Colours indicate different coding regions (grey, 5'UTR and 3'UTR; dark green, signal sequence; light green, Ig-V domain; brown, transmembrane region; red, cytoplasmic tail region).

Appendix Q.




BG8-B12
BG12-B12
BG9-B12
BG13-B12
BG3-B12
BG4-B12
BG6-B12
BG5-B12
BG10-B12
U98-UNK
BG7-B12
BG11-B12
M2-B21
BG0-B12
BG1-B12
BG2-B12


BG8-B12
BG12-B12
BG9-B12
BG13-B12
BG3-B12
BG4-B12
BG6-B12
BG5-B12
BG10-B12
U98-UNK
BG7-B12
BG11-B12
M2-B21
BG0-B12
BG1-B12
BG2-B12



Appendix Q. Nucleotide sequence alignment between two BG transcripts downloaded from GenBank and 14 BG genes of B12 haplotype showing one GenBank sequence is closely related to BG10 and the other one has type 1 cytoplasmic tail. The two BG transcripts, U98-UNK and M2-B21, are downloaded from GenBank with their background information summarized in table 5.1. Names of the genes follow the convention: “BG” and the number of the gene locus for the B12 haplotype. Colours indicate different coding regions (grey, 5’UTR and 3’UTR; dark green, signal sequence; light green, Ig-V domain; brown, transmembrane region; red, cytoplasmic tail region; pink, stop codon).

Appendix R. Summaries of BG mAbs screening by ELISA, Western Blot and red blood cell staining

Lei's label	ELISA	RBC-FAC	WB non-reduced	WB reduced	notes on the flask/tube	numbers (	volume	concentrated	liquid condition
11		x			II-316 +N3 October 90		12		
22		x			20-2D8 18.3.89		70		contanimated
63		x			II-756 +N3 8/10-90		10		
76		x			F21-2.3 21.12.94 +NaN3		10		
94		x			25May 00 Possiblt 2G11-Andy F21-21 SN+0.1%NaN3 7 May 97		10		
98		x			F21-2 TCS 24-3-79 0.1% Axxx		60		
99		x (few)			FIIBu V 50-49 in xxx 4		10	concentrated	contanimated
102		x			F21 21/7 + NaN3 15th May 96+ 20th May 96+ 11 June 96		60		
117		x			I-10F1		25		
132		x			II-326-38 SN+N3 Dec.91		3		
139		x			II-326-41 SN+N3 Dec.91		35		
140		x			II-326-45 SN+N3 Dec.91		5		
141		x			II-326-47 SN+N3 Dec.91		500 ul	concentrated	
142		x			II-326-48 SN+N3 Dec.91		200 ul	concentrated	
143		x			II-326 pool Sept.91 +N3		70		
144		x			II-326-40 SN+N3 Dec.91		50		
145		x			II 326 pool 6/1.92 +N3		70		
146		x			II 326 pool 6/1.92 +N3		60		
147		x			II 326 pool 6/1.92 +N3		70		
203		x			F21-21 SN+N3 9/5-91		1.5		
205		x			F21-2 TC SN March 1999 +0.1%NcN3		80		
206		x			F21-2 TCS +0.01%NaAzide March 1999		150		
207		x			F21-2 TCS 13-4.99 0.1%Nzide		200		
208		x			F21-2 ECS 10.9.99		30		
209		x			I 21-2 SN+N3 14.9.89		10		
210		x			F21-2 TCS 28-3-99 0.1%Azide		200		
224		x			F21-21 30.10 last of TC SN 1 Nov 02		500 ul		
246		x			I-18D2 7.2.		7		
256		x			F12/19a low Ig medium +0.1%NaN3 8.4.98 (110ml)		100		
258		x			F4/21a SN+N3 27.4.98		90		
262		x			F4/21a (E83) TC SN normal media grow 20.4.90 + 0.1%NaN3		40		
263		x			(8F12) F12/19a TC SN 24.3.98 +0.1%NaN3		100		
264		x			F4/21a (EB3) March 98 +0.1%NaN3		120		
268		x			F21-2b SN 6 June 97		100 ul		
274		x			F21-2 mab		1		
275		x			F21-2 clone6 6/4/94		1		yellow color
276		x			37C18 + NaN3 from Ollie		6		
279		x			F21-21 (lots other notes reg. dates and ratio)		50		clouded
281		x			F21-2 TCS 16.3.99 0.1%NaN3		25		

285		x			F21-21 x-x11 19999		5		
x3		x			18-6G2				
158	4,5 (2,3 weak)	x	4,5	W all E7	14-7C11 pool Sept. 91. +N3		155 mL		
229	4,5, (2 weak)	x			14 7C11 SN+N3 28.4.89		28 mL		
110	8,9	x	8,9,12	8,9,12,13	2C-10-2 7/2		100 mL		
137	8,9	x	8,9	8,9	II 301 31/3 concentrated ???pool 22/7 91 +N3		55 mL	concentrated	
148	8,9	x			II 349 14/4 89 concentrated + ????? 18/7 91 added N3	80	50 mL		
153	8,9	x			Fu II 349 4/5.89 +0.5%N3	349	50 mL		
195	8,9,(12,13 weak)	x			I 2C.1D 28/12 87 1/11.90 concentrated		25 mL	concentrated	
199	8,9,(12,13 weak)	x			I-2C10-1 7.2.		110 mL		
200	8,9,(12,13 weak)	x			I2C10 20.11. SN + N3		125 mL		
201	8,9,(12,13 weak)	x	8,9,12	8,12/W all	I 2C10 30/10 concentrated 19/11.90		60 mL	concentrated	
236	8,9,(12,13 weak)	x			I 2C10 +N3 9/4 91		60 mL		
128	1,6,13				I 18-D.11 (28/2.90) concentrated 30/10.90		40 mL	concentrated	
129	1,6,13				I 18D11 30/10 concentrated 16/11.90		25 mL	concentrated	
130	1,6,13		W1,2,3,5,7,8,9	all E 1,3	I 18D11 20/11. concentrated pool 20/11 90		50 mL	concentrated	
8	8,9,12	x	8,9,12	8,11,12/W all E 1,3	II-427 + N3 10/10 90	14	15 mL	No	good
15	8,9,12	x			II-240 + N3 10/10 90	6	0.04 mL	likely	good
55	8,9,13	x	8,12	W0,2,7,8,9,10,12,13	I8D8-3 15.5 pool		25 mL		
131	8,9,13	x	8,9,12/W 11	12/W all E 1,3	I 8D8-33 28/12 89 concentrated pool 6/11 90		50 mL	concentrated	
57	8,9,13	x	8,9,12/W 11	12/W all E 1,3	I8D8-3 pool 1 Feb 90		17 mL		
75	8,9,13	x	8,12/W 9	W 2,8,12/B	J8D8 A 8.5.89 (filtered 24/10/90 due to ?????? Growth by Fiona		15 mL		
56	8,9,13 (1,12 weak)	x	8,9,12	12/W all E 1,3	I8D8-3 pool 1 Feb 90		20 mL		
100	8,9,13	x	8,9,12	12/W all E 1,3	I -2C10 2/4.91		430 mL		
127	8,9,13	x			I - 18C4-3 7.2.		120 mL		
245	8,9,13	x	8,9,12	8,12/W all E 1,3,11	I-18C4-4 7.2.		100 mL		
193	8,9,13	x	8,9,11,12	12/W all E 1,3	I-1A8-2 7.2.		17 mL		
235	8,9,13 (**B**)	x	8,12/W 9,11	12/W all E 1,3	I-1A8-1 7.2.		50 mL		
194	8,9,13	x			I-2E3 7.2.		30 mL		
237	8,9,13	x	8,9,12	8,12/W 2,6,10,13	I-2E3-2 7.2.		60 mL		
189	8,9,13 (**B**)	x	8,12/W 4,9	8,12/W all E 1,3	? F16g 9. (18-6G2) +N3 18/6.90	57	125 mL		
196	8,9,13 (**B**)	x			mAbg 9 +0.1%N3 (18-6G2) 25/6-90		55 mL		
9	8,9,12,13	x			II-409 + N3 10/10 90	13	0.5 mL	likely	good
286	8,9,12,13	x			mAbg 9 (18-6G2)				
289	8,9,12,13	x	8,9,12	all E 1,3	I-8D8 mAbg1 9/4				
91	8,9,12,13(**B**)	x	8,9,11,12	12/W all E 1,3	I-18H6 9.2.		15 mL		
123	1,8,9,12,13	x			I 17A8-1 16/6 89 concentrated 20/11.90		2.5 mL	concentrated	
151	1,8,9,12,13	x	8,9,12/W 10	8,12/W9,13; WW4,5,6	FII8 II 431 25.5.85	431, 81	75 mL		
251	1,8,9,12,13	x	8,12/W 9	12/W all E 1,3	I-18H6-2 7.2.		100 mL		
47	2,5,7,11	x			15 - 3D7 pool 1 Lab 80 95		3.5 mL	No	good

49	2,5,7,11	x	5,7/W 2,11	5,11/W all E 1,3	15 - 3D7 pool 1 Feb 80		9 mL	No	good
176	2,5,7,11,(1 weak)	x	1,5,7,11	2,4,5,11,12,13/W all E1,3	15-3D7 SN +0.1%N3 13/3.90		50 mL		
287	2,5,7,11?	x			15-3D7 pool 2. Feb				
190	4,8,9,13	x			18-6G2 1/6-90 +0.1%N3		170 mL		
202	4,8,9,13	x	4,8,9,12/W 13	8,12/W all E 1,3	18-6G2 pool. May 90 +0.1%N3	77	200 mL		
242	4,8,9,13	x			I-17B8 21/4.89		4 mL		
243	4,8,9,13	x	4,8,9,12/W 11	8,9,12/W all E 1,3	I 17B8-11 SN+N3 30.6.83		50 mL		
10	4,8,9,12	x	8,9,12/W 11	2,8,12/W 3,4,5,6,7,9,10	II-390 + N3 8/10 90	12	13 mL	No	good
37	4,8,9,12,13	x	4,8,9,12,13	8,12/W 4	I 17B8-2 27.5 Pool		30 mL	No	good
38	4,8,9,12,13	x			I 17B8-1 19.5 Pool		22 mL	No	good
39	4,8,9,12,13	x			(I) 17B8 12.5.89		27 mL	No	good
231	1,4,8,9,12,13	x	4,8,9,12	8/W 2,4,5,6,12	I8IE10 SN+N3 28.6.89		80 mL		
107	5	x	4,5,8,9,12	2,5,8/W all E 1	14-8B5 3/1-89		95 mL		
154	5	x	5	2,4,5,6,7,9,10,12/W3,8,11	14-8B5 3/1-89		120 mL		
157	5	x	5/B	All/W 1	148B-5 30/10 concentrated 18/11.90		10 mL	concentrated	
155	5 (**B**)	x	5/W 2,4,7,8,11,12	All	148B5 20.11 SN+N3		100 mL		
156	5 (**B**)	x	5/W 4,6,7,8,11,12	All	14-8B5 28/12.89 concentrated 31/1090		45 mL	concentrated	
77	1 (**B**ackground)				8.5A2 sterile SN 11+-94		1 mL	concentrated	contaminated
93	1 (3 weak) (**B**)		8,9,11,12	W2,4,5,6,7,8,9,12,13	I 19-AJ 27/12 89 concentrated 22/1/90		85 mL	concentrated	
108	1, 12		nothing	nothing	+ II 1066 31.3.	1066	3 mL		contaminated
112	1 (3,7,8,9 weak)		8,9,11,12	7,8,12/W2,4,5,6,9,10,13	II - 874 4/5-89	874	75 mL		
175	2		2,5,8	2	16-3D10 13.6.89 3?		18 mL		
48	2, (1 weak)		2/W12	all	15 - 1B9 19/3-90 + N3	33	9 mL	No	good
53	2, (1 weak)		2	2/B	15 4E3 26/2-90	94	33 mL	No	good
54	2, (1 weak)		2	2/S6,7,12/W0,4,5,8,9,10,11	14-4E3 19/3-90 + N3	92	13 mL	No	good
1					II-690 +N3 October 90		20 mL		good
2					II-568 +N3 October 90		50 mL		
3					II-806 +N3 8/10 - 90		20 mL		
4					II-607 +N3 8/10 -90		20 mL		
5					II-480 +N3 8/10 -90		20 mL		
6					II-477 +N3 8/10-90		20		
7					II-476 +N3 8/10-90		15		
12					II-295 +N3 October 90		17		
13					II-293 +N3 8/10-90		500ul		
14					II-262 +N3 8/10-90		20 mL		
16					II-82 +N3 8/10-90		17		
17					II-68 +N3 8/10-90		15		
18					II-52 +N3 8/10-90		17		
19					20 4D5 28/3		30		yellow color
20					20-4D5 18.3.89		10		yellow color

21					20-4D5 SN 13.3.89		10		
23					20-2B10 28/3		50		contanimated
24					20-2B 10 SN 13.3.89 15III95		15		contanimated
25					I 19C12-2 19.5. pool		15		
26					I19C12-1 15.5. pool		15		
27					I19C12 21/4-89		1	concentrated	
28					I19B5 30/XX Concentrated 21/xx		20	concentrated	contanimated
29					I-18G-12 conentrated (30/10) 16/11-90		2	concentrated	
30					I-18E4-4 15.5. pool		20		
31					I18E4 12.5.83		17		
32					18-6B5 19/3-90 +N3		12		
33					I18E3 30/10 concentrated 21/11 90		17		
34					18-6B5 26/2-90		30		
35					18-2B4 +N3 19/3 90		12		
36					18-2B4 26/2-90		15		
40					I17A8-2 18.5 pool		30		
41					I17A8 8-5/89		35		
42					I-17A8-1 9/5		20		
43					I 17A8 A 8.5.85		25		coulded
44					I-17A-8-1 pool xx 90		5		
45					16-3D3 19/3-90 +N3		8		
46					14 7C 11 pool xxx90		25		
50					14 7C 11 16/9 91 +N3 7/10 91		15		
51					14-8F5-A1 19/3-90 +N3		10		
52					14-8F5-A1 26/2-90		15		
58					A4G3 12.5.89		13		
59					A4G3 12.5.89		13		
60					A4G2 12.5.89		20		
61					II 905 25.5.		20		
62					II-856 N3 October 90.		8		
64					mAbg 2 h 12/6-90		500 ul		
65					II-782 16/9 91 +N3 7/10 91		2		
66					II 856 16/9 91 +N3 7/10 91		5		
67					II-785 +N3 October 90		6		
68					A4G 2 12.5.89		15		
69					A1C6 12.5.89		1		
70					mAbg 2 g 12/6-90		35		
71					mAbg 2 f 12/6-90		35		
72					mAbg 2 e 12/6-90		30		
73					mAbg 2 d 12/6-90		30		

74					mAbg 2 c 12/6-90		30		
78					I-18F6		50		
79					I-17G2 7.2.		50		
80					I18E-3(23/12.83) concentraded 31/10.90		10		
81					I-18E3 7.2.		40		
82					I 18E4-43 SN+N3 30.6.89		15		
83					I-18E4-A +N3 21/4-89 (had xxxx)		1		
84					I18F6 30/10 concentrated 21/11 90		10	concentrated	
85					I18F6 28/12 89 concentrated 16/11.90		15	concentrated	
86					I-18G9B 21/4-89		1		
87					I 18G-12 27/12 89 concentrated 16/11.90		7	concentrated	
88					I-19A2		20		
89					I-18G12		20		
90					I-19-B5 (28/12.90) concentrated 30/10.90		7	concentrated	
92					I-19A3		40		
95					(I) 19A-3 concentrated 30/10		3		
96					I-19B5-2 7.2.		7		
97					2G11 TC SN +0.1%NaN3 14/11/96 to		50		
95					(I) 19A-3 concentrated 30/10		3		
96					I-19B5-2 7.2.		7		
97					2G11 TC SN +0.1%NaN3 14/11/96 to		50		
101					IV-9D3 SN+N3 March 92		7		
103					FIIB uN 47-44 1/5 13 VI 94		20	concentrated	contanimated
104					4E12 SN March 92 +N3 Jan		7		
105					15-2D5 3/1-89		30		
106					I148F5 20.11. SN+N3		25		
109					mix of F11B47-44 & 50-49 in selection medium 18April 94 No NaN3		75		contanimated
111					923		10		
113					II-856 pool Sept. 91 +N3		50		
114					II-782 pool Sept.91 +N3		25		
115					xxx(Fus?) II 1019 25.5.89		25		
116					II-1066 +N3 14/4.89		5		
118					I17A3 30/10 concnetrated 20/11 90		20	concentrated	
119					I-17A3-5 7.2		15		
120					I 17A3 (27/12.89) concentrated 2/11 90		13	concentrated	
121					I-17A4 7.2		40		
122					I-18D2		15		
124					I-17A8-11 SN+N3 30.6.89		5		
125					I-17A8-B 21/4.89		200 ul	concentrated	clouded
126					I-17B7-2 7.2.		15		

133					I 19C12-11 SN+N3 10.6.89		10		
134					Fus II 141 25.3.89		10		
135					Fus II 190 25.5.89		20		
136					Fus II 235 25.5.89		7		
138					II352 31/3 concentrated +filltered 22/7.91 +N3		1		
149					II 352 4/5.89 concentrated+filtered 22/7.91 +N3		15		
150					Fu II 562 31/3.89 +0.5N3		10		
152					F II 636 31.3		20		
159					14-7C11 pool 2.Feb 91		20		
160					148F5 SN+N3 26.6.89		10		
161					148F5 30/10 concentrated 20/1.90		10	concentrated	
162					14 8F3 28/12.89 concentrated 6/11.90		10	concentrated	
163					mAbg 2a. +0.1%N3 25/6-90		10		
164					mAbg 2b. +0.1%N3 25/6.90		12		
165					147E2 SN+N3 26.6.89		8		
166					147E2 28/12 89 concentrated 21/11.90		30	concentrated	
167					14-7E2 C SN+N3 31.8.1989		50		
168					14 7E2-2 13/11 concentrated 18/11.90		15	concentrated	
169					α CD2 2-4-8 SN+N3 11.9.89		30		
170					2-6-1 SN+N3 11.9.89		35		
171					11-A-9 SN+N3 14.9.89		30		
172					11-38-6 SN+N3 14.9.89		40		
173					11-39-4 SN+N3 11.9.89		45		
174					16-7 B9 SN+N3 23.6.89		20		
177					14-8F5-A1 pool 1.Feb 90		200 ul	concentrated	
178					14-7E2-C9 pool 12.10.89		15		
179					B IV-7C11 +N3 Feb 92		150		
180					IV-9D3 +N3 Feb 92		150		orange color
181					2G11 +N3 17/4.89		100		contanminated
182					2G11 +N3 7/4-89		100		contanminated
183					IV-4E12 SN+N3 Martch 92		200		orange color
184					IV 4E12 pool 7.4.92 +N3		100		
185					2G11 sup+0.1%N3 May 90		60		
186					IV-9D3 +N3 Feb 90		250		
187					2G11 sup+0.1%N3 5/3 90		200		
188					IV-7C11 +N3 Feb 92		100		
191					15-2D5 3/1-89		50		
192					18-6G2 'old' Sept-90 +N3		60		
197					18-6G2 'OLD' +N3 23/8-90		30		
198					18-6G2 'old' +N3 24/8-90		40		

204					xF3 SN+N3 23.6.85		20		
211					2G11 TC SN 7/11/96 to 14/11/96 + 0.1% NaN3		200		
212					2G11 + 0.1%NaN3 22/11/96 + 22/11/96		200		
213					2G11 + 0.1%NaN3 5/12/96 11/12/96		170		
214					2G11 SN +0.1%NaN3 23 April'97		200		
215					A1C6 22.5		25		
216					A4G3 22.5		10		
217					Fu 2-3-5 SN+N3 14.9.89		100		
218					Fu 9-8 SN+N3 14.9.88		50		
219					11-G-2 SN+N3 11.9.89		25		orange color
220					A4G2 22.5		30		
221					11G2 Ollie's Bu-1b TC SN from big bottle 28 Feb 95		15		
222					CB3 SN+0.1%N3 2/11-1992		40		
223					L22 SN+N3 11/9/89 filltered on 24/10/94 due to fungi growth		20		yellow color
225					I-19C12 21/8-89		5		yellow color
226					F II 52 31.3.		30		
227					II 64 31.3		40		
228					Fus II 139 25.5.89		50		
230					16-8D7 (sTCR) Sept 90 +N3		70		
232					I86G2.89-F10-1 SN+N3 23.6.85		25		
233					I-19B5 7.2.		50		
234					21-1A6 SN+N3 11.9.89		50		
238					I-10F1		10		
239					I-17A3 7.2.		20		
240					I-17A4 7.2.		50		
241					I-17B7 7.2		30		
244					I-17G2 7.2.		40		
247					I18E3-21 7.2.		5		
248					I-18F6 7.2.		35		
249					I-18G9-A +N3 21/4-89		3		
250					I-18G12 7.2.		50		
252					I-19A2 7.2.		20		
253					I-19A3 7.2.		50		
254					I-19A7 7.2.		15		
255					I-19B2 7.2.		40		
257					F12/19a TC SN (cell in low Ig media) April/98 +0.1%NaN3		100		
259					F4/2/9 grow in low Ig media 29.4.98		90		
260					F4/21b (KA4) TC SN grow in low Ig media 11.5.98		200		
261					F4/21b grow in low Ig media (KA4) 29.4.98 +0.1%NaN3		170		
265					F4/21b (KA4) 14.5.98 +0.1%NaN3		15		

266					6.3.98 + 9.3.98 F12/19a *RF12) +0.1%NaN3		110		
267					F4/21b (KA4) TC SN March 98 +0.1%NaN3		110		
269					2G11 + 0.1%NaN3 30-5-90 filltered 13/2/98		1.5		
270					2G11 TC SN 29/10/96 FJ +0.1%NaN3		20		
271					3-292		3		
272					2G11 5/3-90		8		
273					16-8D7C Sept 90 +N3		20		
277					16-8D7 Sep 90		4		
278					II-856 27/8-91		100 ul	concentrated	
280					29x SN 6 Jan 97		4		
282					2G11 TC SN 11.3.98		20	concentrated	contaminated
283					2G11 TC SN 8.3.98		100		
284					CT3 supernatant 31.8.95 contain 0.1%NaN3		25		
288					I-19A5 October 89 concentrated 19/11.90		1		
290					I-8D8-A 26/4-89		500 ul		
291					m17bg 2 a (14-7C11) SN+N3 12/6-90		20		
X1					I-808				
x2					mAbg2 (14-7C11)				
x4					I-17A8-1				
x5					5-6CB				
x6					5-7D7				

Reference

Abeler-Dörner, L., Swamy, M., Williams, G., Hayday, A. C., & Bas, A. (2012). Butyrophilins: an emerging family of immune regulators. *Trends in Immunology*, *33*(1), 34–41.

Abplanalp, H., Sato, K., Napolitano, D., & Reid, J. (1992). Reproductive performance of inbred congenic leghorns carrying different haplotypes for the major histocompatibility complex. *Poultry Science*, *71*(1), 9–17.

Akinc, A., Thomas, M., Klibanov, A. M., & Langer, R. (2005). Exploring polyethylenimine-mediated DNA transfection and the proton sponge hypothesis. *The Journal of Gene Medicine*, *7*(5), 657–663.

Ammann, J. U., Cooke, A., & Trowsdale, J. (2013). Butyrophilin Btn2a2 inhibits TCR activation and phosphatidylinositol 3-kinase/Akt pathway signaling and induces Foxp3 expression in T lymphocytes. *Journal of Immunology (Baltimore, Md. : 1950)*, *190*(10), 5030–6.

Ammann, J. U., Jahnke, M., Dyson, M. R., Kaufman, J., & Trowsdale, J. (2012). Detection of weak receptor–ligand interactions using IgM and J-chain-based fusion proteins. *European Journal of Immunology*, *42*(5), 1354–1356.

Ampuero, J., del Campo, J. A., Rojas, L., Garcia-Lozano, R. J., Buti, M., Sola, R., … Romero-Gomez, M. (2015). Fine-mapping butyrophilin family genes revealed several polymorphisms influencing viral genotype selection in hepatitis C infection. *Genes and Immunity*, *16*(5), 297–300.

Aronsson, C., Dånmark, S., Zhou, F., Öberg, P., Enander, K., Su, H., & Aili, D. (2015). Self-sorting heterodimeric coiled coil peptides with defined and tuneable self-assembly properties. *Scientific Reports*, *5*, 14063.

Bacon, L. D., Hunt, H. D., & Cheng, H. H. (2000). A Review of the Development of Chicken Lines to Resolve Genes Determining Resistance to Diseases. *Poultry Science*, *79*(8), 1082–1093.

Barbee, S. D., Woodward, M. J., Turchinovich, G., Mention, J.-J., Lewis, J. M., Boyden, L. M., … Hayday, A. C. (2011). Skint-1 is a highly specific, unique selecting component for epidermal T cells. *Proceedings of the National Academy of Sciences*, *108*(8), 3330–3335.

Bas, A., Swamy, M., Abeler-dörner, L., Williams, G., Pang, D. J., & Barbee, S. D. (2011). Butyrophilin-like 1 encodes an enterocyte protein that selectively regulates functional interactions with T lymphocytes. *Proceedings of the National Academy of Sciences*, 108(11), 4376–4381.

Beck, S., & Trowsdale, J. (2000). The human major histocompatibility complex: lessons from the DNA sequence. *Annual Review of Genomics and Human Genetics*, *1*(1), 117–137.

Benyamine, A., Loncle, C., Foucher, E., Blazquez, J.-L., Castanier, C., Chrétien, A.-S., … Olive, D. (2018). BTN3A is a prognosis marker and a promising target for Vγ9Vδ2 T cells based-immunotherapy in pancreatic ductal adenocarcinoma (PDAC). *OncoImmunology*, *7*(1), e1372080.

Bernard, C. C., Johns, T. G., Slavin, A., Ichikawa, M., Ewing, C., Liu, J., & Bettadapura, J. (1997). Myelin oligodendrocyte glycoprotein: a novel candidate autoantigen in multiple sclerosis. *Journal of Molecular Medicine (Berlin, Germany)*, *75*(2), 77–88.

Bikle, D. D., Munson, S., Morrison, N., & Eisman, J. (1993). Zipper protein, a newly described tropomyosin-like protein of the intestinal brush border. *Journal of Biological Chemistry*, *268*(1), 620–626.

Bikle, D. D., Munson, S., & Komuves, L. (1996). Zipper protein, a B-G protein with the ability to regulate actin/myosin 1 interactions in the intestinal brush border. *Journal of Biological Chemistry*, *271*(15), 9075–9083.

Bourlet, Y., Béhar, G., Guillemot, F., Fréchin, N., Billault, A., Chaussé, A. M., … Auffray, C. (1988). Isolation of chicken major histocompatibility complex class II (B-L) beta chain sequences: comparison with mammalian beta chains and expression in lymphoid organs. *The EMBO Journal*, *7*(4), 1031–9.

Boyle, L. H., Hermann, C., Boname, J. M., Porter, K. M., Patel, P. A., Burr, M. L., … Trowsdale, J. (2013). Tapasin-related protein TAPBPR is an additional component of the MHC class I presentation pathway. *Proceedings of the National Academy of Sciences*, *110*(9), 3465–3470.

Boyden, L. M., Lewis, J. M., Barbee, S. D., Bas, A., Girardi, M., Hayday, A. C., … Lifton, R. P. (2008). Skint1, the prototype of a newly identified immunoglobulin superfamily gene cluster, positively selects epidermal γδ T cells. *Nature Genetics*, *40*(5), 656–662.

Briles, W. E., Bumstead, N., Ewert, D. L., Gilmour, D. G., Gogusev, J., Hala, K., … Pink, J. R. L. (1982). Nomenclature for chicken major histocompatibility (B) complex. *Immunogenetics*, *15*(5), 441–447.

Briles, W. E., & Briles, R. W. (1977). Some recent recombinants at the B locus. In *Avian Immunology* (pp. 221–225). Springer.

Briles, W. E., & Briles, R. W. (1980). A search for `duplicate' recombinants between the BF and BG regions of the chicken B complex. *Anim Blood Groups Biochem Genet*, *11*(Suppl 1), 38–39.

Briles, W. E., McGibbon, W. H., & Irwin, M. R. (1950). On multiple alleles effecting cellular antigens in the chicken. *Genetics*, *35*(6), 633–652.

Butter, C., Staines, K., van Hateren, A., Davison, T. F., & Kaufman, J. (2013). The peptide motif of the single dominantly expressed class I molecule of the chicken MHC can explain the response to a molecular defined vaccine of infectious bursal disease virus (IBDV). *Immunogenetics*, *65*(8), 609–618.

Campbell, K. S., & Purdy, A. K. (2011). Structure/function of human killer cell immunoglobulin-like receptors: lessons from polymorphisms, evolution, crystal structures and mutations. *Immunology*, *132*(3), 315–325.

Ceeraz, S., Nowak, E. C., & Noelle, R. J. (2013). B7 family checkpoint regulators in immune regulation and disease. *Trends in Immunology*, *34*(11), 556–563.

Chan, A. (2014). *The genetics of vaccine responses in experimental lines of chickens*. University of Cambridge.

Chang, C. (2004). *Human decidual NK cell receptors and their functions.* University of Cambridge.

Charleston, B., Gate, J. J., Aitken, I. A., Stephan, B., & Froyman, R. (1998). Comparison of the efficacies of three fluoroquinolone antimicrobial agents, given as continuous or pulsed-water medication, against Escherichia coli infection in chickens. *Antimicrobial Agents and Chemotherapy*, *42*(1), 83–87.

Chattaway, J. (2013). *Characterising the polymorphic BG gene family.* University of Cambridge.

Chattaway, J., Ramirez-Valdez, R. A., Chappell, P. E., Caesar, J. J. E., Lea, S. M., & Kaufman, J. (2016). Different modes of variation for each BG lineage suggest different functions. *Open Biology*, *6*(9).

Chaussé, A.-M., Coudert, F., Dambrine, G., Guillemot, F., Miller, M. M., & Auffray, C. (1989). Molecular genotyping of four chicken B-complex haplotypes with B-Lβ, BF, and BG probes. *Immunogenetics*, *29*(2), 127–130.

Chaves, L. D., Faile, G. M., Krueth, S. B., Hendrickson, J. a, & Reed, K. M. (2010). Haplotype variation, recombination, and gene conversion within the turkey MHC-B locus. *Immunogenetics*, *62*(7), 465–477.

Chen, L., Fakiola, M., Staines, K., Butter, C., & Kaufman, J. (2018). Functional alleles of chicken BG genes, members of the butyrophilin gene family, in peripheral T cells. *Frontiers in Immunology*, *9*, 930.

Clayton, K. L., Douglas-Vail, M. B., Rahman, A. K. M. N., Medcalf, K. E., Xie, I. Y., Chew, G. M., … Ostrowski, M. A. (2015). Soluble T cell immunoglobulin mucin domain 3 is shed from CD8(+) T cells by the sheddase ADAM10, is increased in plasma during untreated HIV infection, and correlates with HIV disease progression. *Journal of Virology*, *89*(7), 3723–3736.

Collins, W. M., Briles, W. E., Zsigray, R. M., Dunlop, W. R., Corbett, A. C., Clark, K. K., … McGrail, T. P. (1977). The B locus (MHC) in the chicken: association with the fate of RSV-induced tumors. *Immunogenetics*, *5*(1), 333–343.

Cotter, P. F., Taylor R. L., J., & Abplanalp, H. (1998). B-complex associated immunity to Salmonella enteritidis challenge in congenic chickens. *Poultry Science*, *77*(12), 1846–1851.

Dawes, M. E., Griggs, L. M., Collisson, E. W., Briles, W. E., & Drechsler, Y. (2014). Dramatic differences in the response of macrophages from B2 and B19 MHC-defined haplotypes to interferon gamma and polyinosinic:polycytidylic acid stimulation. *Poultry Science*, *93*(4), 830–838.

Cheng, J., Zhou, T., Liu, C., Shapiro, J. P., Brauer, M. J., Kiefer, M. C., … Mountz, J. D. (1994). Protection from fas-mediated apoptosis by a soluble form of the fas molecule. *Science*, *263*(5154), 1759–1762.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, *5*, 113.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.

Emara, M. G., Kim, H., Zhu, J., Lapierre, R. R., Lakshmanan, N., & Lillehojt, H. S. (2002). Genetic diversity at the major histocompatibility complex (B) and microsatellite loci in three commercial broiler pure lines. *Poultry Science*, *81*(11), 1609–1617.

Ewald, S. J., & Livant, E. J. (2004). Distinctive polymorphism of chicken B-FI (major histocompatibility complex class I) molecules. *Poultry Science*, *83*(4), 600–605.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, *39*(4), 783–791.

Gilmour, D. G., Collins, W. M., Fredericksen, T. L., Auclair, B., Clark, K. K., & Urban, W. E. (1983). Influence of non-MHC T lymphocyte alloantigens on regression of rous sarcomas in the chicken. *Immunogenetics*, *17*(1), 43–54.

Goto, R., Miyada, C. G., Young, S., Wallace, R. B., Abplanalp, H., Bloom, S. E., … Miller, M. M. (1988). Isolation of a cDNA clone from the BG subregion of the chicken histocompatibility (B) complex. *Immunogenetics*, *27*, 102–109.

Goto, R. M., Wang, Y., Taylor, R. L., Wakenell, P. S., Hosomichi, K., Shiina, T., … Miller, M. M. (2009). BG1 has a major role in MHC-linked resistance to malignant lymphoma in the chicken. *Proceedings of the National Academy of Sciences*, *106*(39), 16740–16745.

Gu, S., Sachleben, J. R., Boughter, C. T., Nawrocka, W. I., Borowska, M. T., Tarrasch, J. T., … Adams, E. J. (2017). Phosphoantigen-induced conformational change of butyrophilin 3A1 (BTN3A1) and its implication on Vγ9Vδ2 T cell activation. *Proceedings of the National Academy of Sciences*, *114*(35), E7311–E7320.

Guillemot, F., Billault, a, Pourquié, O., Béhar, G., Chaussé, a M., Zoorob, R., … Auffray, C. (1988). A molecular map of the chicken major histocompatibility complex: the class II beta genes are closely linked to the class I genes and the nucleolar organizer. *The EMBO Journal*, *7*(9), 2775–2785.

Hála, K., Vilhelmova, M., Schulmannova, J., & Plach`y, J. (1979). A new recombinant allele in the B complex of the chicken [proceedings]. *Folia Biologica*, *25*(5), 323.

Hála, K., Chaussé, A.-M., Bourlet, Y., Lassila, O., Hasler, V., & Auffray, C. (1988). Attempt to detect recombination between BF and BL genes within the chicken B complex by serological typing, in vitro MLR, and RFLP analyses. *Immunogenetics*, *28*(6), 433–438.

Hála, K., Boyd, R., & Wick, G. (1981a). Chicken major histocompatibility complex and disease. *Scandinavian Journal of Immunology*, *14*(6), 607–616.

Hála, K., Plachý, J., & Schulmannová, J. (1981b). Role of the BG-region antigen in the humoral immune response to the BF-region antigen of chicken MHC. *Immunogenetics*, 393–401.

Hála, K., Vilhelmová, M., & Hartmanová, J. (1976). Probable crossing-over in the B blood group system of chickens. *Immunogenetics*, *3*(1), 97–103.

Harbers, M. (2008). The current status of cDNA cloning. *Genomics*, *91*(3), 232–242.

Harly, C., Guillaume, Y., Nedellec, S., Peigné, C.-M., Mönkkönen, H., Mönkkönen, J., … Scotet, E. (2012). Key implication of CD277/butyrophilin-3 (BTN3A) in cellular stress sensing by a major human γδ T-cell subset. *Blood*, *120*(11), 2269–2279.

Henry, J., Miller, M. M., & Pontarotti, P. (1999). Structure and evolution of the extended B7 family. *Immunology Today*, *20*(6), 285–288.

Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., Ponting, C. P., … Delany, M. E. (2004). Sequencing and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, *432*(7018), 695–716.

Hiramatsu, M., Oguri, M., Kato, K., Yoshida, T., Fujimaki, T., Horibe, H., … Yamada, Y. (2011). Association of a polymorphism of BTN2A1 with Type 2 diabetes mellitus in Japanese individuals. *Diabetic Medicine*, *28*(11), 1381–1387.

Hsueh, K.-C., Lin, Y.-J., Chang, J.-S., Wan, L., & Tsai, F.-J. (2010). BTNL2 gene polymorphisms may be associated with susceptibility to Kawasaki disease and formation of coronary artery lesions in Taiwanese children. *European Journal of Pediatrics*, *169*(6), 713–719.

Hosomichi, K., Shiina, T., Suzuki, S., Tanaka, M., Shimizu, S., Iwamoto, S., … Hanzawa, K. (2006). The major histocompatibility complex (Mhc) class IIB region has greater genomic structural flexibility and diversity in the quail than the chicken. *BMC Genomics*, *7*, 322.

Hosomichi, K., Miller, M. M., Goto, R. M., Wang, Y., Suzuki, S., Kulski, J. K., … Shiina, T. (2008). Contribution of mutation, recombination, and gene conversion to chicken MHC-B haplotype diversity. *Journal of Immunology (Baltimore, Md. : 1950)*, *181*(5), 3393–3399.

Hung, J.-H., & Weng, Z. (2016). Sequence alignment and homology search: Table 1. *Cold Spring Harbor Protocols*, *2016*(11), pdb.top093070.

Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., Ponting, C. P., … Delany, M. E. (2004). Sequencing and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, *432*(7018), 695–716.

Iglesias, G. M., Huguet, M. J., Goto, R. M., Miquel, M. C., & Miller, M. M. (2007). Report of new alleles at BG loci in Camperos chickens. *BAG. Journal of Basic and Applied Genetics*, *18*(1), 29–30.

Iglesias, G. M., Soria, L. A., Goto, R. M., Jar, A. M., Miquel, M. C., Lopez, O. J., & Miller, M. M. (2003). Genotypic variability at the major histocompatibility complex (B and Rfp-Y) in Camperos broiler chickens. *Animal Genetics*, *34*(2), 88–95.

Jacob, J. P., Milne, S., Beck, S., & Kaufman, J. (2000). *The major and a minor class II B-chain (B-LB) gene flank the Tapasin gene in the B-F/B-L region of the chicken major histocompatibility complex*. *Immunogenetics* (Vol. 51).

Jean, A. (2001). Brain stem control of swallowing: neuronal network and cellular mechanisms. *Physiological Reviews*, *81*(2), 929–969.

Jeannin, P., Magistrelli, G., Aubry, J.-P., Caron, G., Gauchat, J.-F., Renno, T., … Delneste, Y. (2000). Soluble CD86 is a costimulatory molecule for human T lymphocytes. *Immunity*, *13*(3), 303–312.

Joiner, K. S., Hoerr, F. J., van Santen, E., & Ewald, S. J. (2005). The avian major histocompatibility complex influences bacterial skeletal disease in broiler breeder chickens. *Veterinary Pathology*, *42*(3), 275–281.

Juul-Madsen, H. R., Dalgaard, T. S., & Afanassieff, M. (2008). Molecular characterization of major and minor MHC class I and II genes in B21-like haplotypes in chickens. *Animal Genetics*, *31*(4), 252–261.

Kaufman, J., Salomonsen, J., Skjodt, K., Thorpe, D., Skjødt, K., & Thorpe, D. (1990). Size polymorphism of chicken major histocompatibility complex-encoded B-G molecules is due to length variation in the cytoplasmic heptad repeat region. *Proceedings of the National Academy of Sciences*, *87*(21), 8277–8281.

Kaufman, J., Salomonsen, J., & Skjødt, K. (1989). BG cDNA clones have multiple small repeats and hybridize to both chicken MHC regions. *Immunogenetics*, *30*(6), 440–451.

Kaufman, J., Milne, S., Göbel, T. W., Walker, B. a, Jacob, J. P., Auffray, C., … Beck, S. (1999). The chicken B locus is a minimal essential major histocompatibility complex. *Nature*, *401*(6756), 923–925.

Kaufman, J. (2008). Chapter 8 - The Avian MHC. In F. Davison, B. Kaspers, & K. A. Schat (Eds) - *Avian Immunology, first edition* (pp. 161–183). London: Academic Press.

Kaufman, J. (2014). Chapter 8 - The Avian MHC. In K. A. Schat, B. Kaspers, & P. Kaiser (Eds) - *Avian Immunology, second edition* (pp. 149–167). Boston: Academic Press.

Kaufman, J. (2015). Co-evolution with chicken class I genes. *Immunological Reviews*, *267*(1), 56–71.

Kaufman, J. (2018a). Generalists and Specialists: A new view of how MHC class I molecules fight infectious pathogens. *Trends in Immunology*, *39*(5), 367–379.

Kaufman, J. (2018b). Unfinished Business: Evolution of the MHC and the adaptive immune system of jawed vertebrates. *Annual Review of Immunology*, *36*(1), 383–409.

Kaufman, J., Andersen, R., Avila, D., Engberg, J., Lambris, J., Salomonsen, J., … Skjødt, K. (1992). Different features of the MHC class I heterodimer have evolved at different rates. Chicken B-F and beta 2-microglobulin sequences reveal invariant surface residues. *Journal of Immunology (Baltimore, Md. : 1950)*, *148*(5), 1532–1546.

Kaufman, J., Skjødt, K., & Salomonsen, J. (1991). The BG multigene family of the chicken major histocompatibility complex. *Critical Reviews in Immunology*, *11*(2), 113.

Kaufman, J., Völk, H., & Wallny, H.-J. (1995). A "minimal essential Mhc" and an "unrecognized Mhc": two extremes in selection for polymorphism. *Immunological Reviews*, *143*(1), 63–88.

Kim, T., Hunt, H. D., Parcells, M. S., van Santen, V., & Ewald, S. J. (2018). Two class I genes of the chicken MHC have different functions: BF1 is recognized by NK cells while BF2 is recognized by CTLs. *Immunogenetics*, *70*(9), 599–611.

Kline, K., Briles, W. E., Bacon, L., & Sanders, B. G. (1988). Characterization of two distinct disulfide-linked BG molecules in the chicken. *Journal of Heredity*, *79*(4), 249–256.

Koch, C., Skjosdt, K., Toivanen, A., & Toivanen, P. (1983). New recombinants within the MHC (B-complex) of the chicken. *Tissue Antigens*, *21*(2), 129–137.

Koch, M., Camp, S., Collen, T., Avila, D., Salomonsen, J., Wallny, H.-J., … Kaufman, J. (2007). Structures of an MHC class I molecule from B21 chickens illustrate promiscuous peptide binding. *Immunity*, *27*(6), 885–899.

Lamont, S. J., Bolin, C., & Cheville, N. (1987). Genetic resistance to fowl cholera is linked to the major histocompatibility complex. *Immunogenetics*, *25*(5), 284–289.

Lanier, L. L. (2009). DAP10- and DAP12-associated receptors in innate immunity. *Immunological Reviews*, *227*(1), 150–160.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., McWilliam, H., … Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*(21), 2947–2948.

Li, L., Johnson, L. W., & Ewald, S. J. (1997). Molecular characterization of major histocompatibility complex (B) haplotypes in broiler chickens. *Animal Genetics*, *28*(4), 258–267.

Li, L., Johnson, L. W., Livant, E. J., & Ewald, S. J. (1999). The MHC of a broiler chicken line: serology, B-G genotypes, and B-F/B-LB sequences. *Immunogenetics*, *49*(3), 215–224.

Linsley, P. S., Peach, R., Gladstone, P., & Bajorath, J. (1994). Extending the B7 (CD80) gene family. *Protein Science : A Publication of the Protein Society*, *3*(8), 1341–3.

Livant, E. J., Zheng, D., Johnson, L. W., Shi, W., & Ewald, S. J. (2001). Three new MHC haplotypes in broiler breeder chickens. *Animal Genetics*, *32*(3), 123–131.

Livant, E. J., Brigati, J. R., & Ewald, S. J. (2004). Diversity and locus specificity of chicken MHC B class I sequences. *Animal Genetics*, *35*(1), 18–27.

Longenecker, B. M., Mosmann, T. R., & Shiozawa, C. (1979). A strong, preferential response of mice to polymorphic antigenic determinants of the chicken MHC, analyzed with mouse hybridoma (monoclonal) antibodies. *Immunogenetics*, *9*(1), 137–147.

Lobner, E., Humm, A.-S., Göritzer, K., Mlynek, G., Puchinger, M. G., Hasenhindl, C., … Obinger, C. (2017). Fcab-HER2 Interaction: a Ménage à Trois. Lessons from X-Ray and Solution Studies. *Structure*, *25*(6), 878–889.e5.

Mailund, T., Brodal, G. S., Fagerberg, R., Pedersen, C. N. S., & Phillips, D. (2006). Recrafting the neighbor-joining method. *BMC Bioinformatics*, *7*, 29.

Maruoka, T., Tanabe, H., Chiba, M., & Kasahara, M. (2005). Chicken CD1 genes are located in the MHC: CD1 and endothelial protein C receptor genes constitute a distinct subfamily of class-I-like genes that predates the emergence of mammals. *Immunogenetics*, *57*(8), 590–600.

McBride, R. A., & Schierman, L. W. (1970). Hapten-carrier relationships of isoantigens: a model for immunological maturation based on the conversion of haptens to carriers by antibody. *The Journal of Experimental Medicine*, *131*(2), 377–390.

McBride, R. A., & Schierman, L. W. (1971). Passive antibody and the immune response: factors which determine enhancement and suppression. *The Journal of Experimental Medicine*, *134*(4), 833–845.

Miller, M. M., Goto, R., Young, S., Chirivella, J., Hawke, D., & Miyada, C. G. (1991). Immunoglobulin variable-region-like domains of diverse sequence within the major histocompatibility complex of the chicken. *Proceedings of the National Academy of Sciences*, *88*(10), 4377–4381.

Miller, M. M., Goto, R., & Abplanalp, H. (1984). Analysis of the BG antigens of the chicken MHC by two-dimensional gel electrophoresis. *Immunogenetics*, 373–385.

Miller, M. M., Goto, R., Young, S., Liu, J., & Hardy, J. (1990). Antigens similar to major histocompatibility complex BG are expressed in the intestinal epithelium in the chicken. *Immunogenetics*, *32*(1), 45–50.

Miller, M. M. (1990). The Major Histocompatibility Complex of the Chicken. In G. W. Warr & N. Cohen (Eds.), *Phylogenesis of immune functions* (p. 161). Boca Raton : CRC Press.

Miller, M. M., Goto, R., Zoorob, R., Auffray, C., & Elwood Briles, W. (1994). Regions of homology shared by Rfp-Y and major histocompatibility B complex genes. *Immunogenetics*, *39*(1), 71–73.

Miller, M. M., Goto, R., Bernot, A., Zoorob, R., Auffray, C., Bumstead, N., & Briles, W. E. (1994). Two Mhc class I and two Mhc class II genes map to the chicken Rfp-Y system outside the B complex. *Proceedings of the National Academy of Sciences*, *91*(10), 4397–4401.

Miller, M. M., Bacon, L. D., Hala, K., Hunt, H. D., Ewald, S. J., Kaufman, J., … Briles, W. E. (2004). 2004 Nomenclature for the chicken major histocompatibility (B and Y) complex. *Immunogenetics*, *56*(4), 261–279.

Miller, M. M., Abplanalp, H., & Goto, R. (1988). Genotyping chickens for the BG subregion of the major histocompatibility complex using restriction fragment length polymorphisms. *Immunogenetics*, *28*(5), 374-379.

Miller, M. M., Goto, R., & Clark, S. D. (1982). Structural characterization of developmentally expressed antigenic markers on chicken erythrocytes using monoclonal antibodies. *Developmental Biology*, *94*(2), 400–414.

Miller, M. M., & Taylor, R. L. (2016). Brief review of the chicken Major Histocompatibility Complex: the genes, their distribution on chromosome 16, and their contributions to disease resistance. *Poultry Science*, *95*(2), 375–392.

Møller, L. B., Kaufman, J., Verland, S., Salomonsen, J., Avila, D., Lambris, J. D., & Skjødt, K. (1991). Variations in the cytoplasmic region account for the heterogeneity of the chicken MHC class I (B-F) molecules. *Immunogenetics*, *34*(2), 110–120.

Moon, R. T., Ngai, J., Wold, B. J., & Lazarides, E. (1985). Tissue-specific expression of distinct spectrin and ankyrin transcripts in erythroid and nonerythroid cells. *Journal of Cell Biology*, *100*(1), 152–160.

Nei, M., Kumar, S., & Takahashi, K. (1998). The optimization principle in phylogenetic analysis tends to give incorrect topologies when the number of nucleotides or amino acids used is small. *Proceedings of the National Academy of Sciences*, *95*(21), 12390–12397.

O'Neill, A. M., Livant, E. J., & Ewald, S. J. (2009). The chicken BF1 (classical MHC class I) gene shows evidence of selection for diversity in expression and in promoter and signal peptide regions. *Immunogenetics*, *61*(4), 289–302.

Norup, L. R., Dalgaard, T. S., Pedersen, A. R., & Juul-Madsen, H. R. (2011). Assessment of newcastle disease–specific T cell proliferation in different inbred MHC chicken lines. *Scandinavian Journal of Immunology*, *74*(1), 23–30.

Oaks, M. K., & Hallett, K. M. (2000). Cutting Edge: A soluble form of CTLA-4 in patients with autoimmune thyroid disease. *The Journal of Immunology*, *164*(10), 5015 LP-5018.

Parker, A. (2013). *Classical and non-classical major histocompatibility complex class II genes in the chicken.* University of Cambridge.

Parker, A., & Kaufman, J. (2017). What chickens might tell us about the MHC class II system. *Current Opinion in Immunology*, *46*, 23–29.

Pickel, J. M., Chen, C. H., & Cooper, M. D. (1990). An avian B-lymphocyte protein associated with β2-microglobulin. *Immunogenetics*, *32*(1), 1–7.

Pink, J. R. L., Droege, W., Hala, K., Miggiano, V. C., & Ziegler, A. (1977). A three-locus model for the chicken major histocompatibility complex. *Immunogenetics*, *5*(1), 203–216.

Régnier, V., Novelli, J., Fukagawa, T., Vagnarelli, P., & Brown, W. (2003). Characterization of chicken CENP-A and comparative sequence analysis of vertebrate centromere-specific histone H3-like proteins. *Gene*, *316*, 39–46.

Rhodes, D. A., Chen, H.-C., Price, A. J., Keeble, A. H., Davey, M. S., James, L. C., … Trowsdale, J. (2015). Activation of human γδ T cells by cytosolic interactions of BTN3A1 with soluble phosphoantigens and the cytoskeletal adaptor periplakin. *Journal of Immunology (Baltimore, Md. : 1950)*, *194*(5), 2390–8.

Rhodes, D. A., Reith, W., & Trowsdale, J. (2016). Regulation of immunity by butyrophilins. *Annual Review of Immunology*, *34*(1), annurev-immunol-041015-055435.

Ruby, T., Bed'Hom, B., Wittzell, H., Morin, V., Oudin, A., & Zoorob, R. (2005). Characterisation of a cluster of TRIM-B30.2 genes in the chicken MHC B locus. *Immunogenetics*, *57*(1–2), 116–128.

Saitou, N., & Nei, M. (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evo*, *4*(4), 406–425.

Salim, M., Knowles, T. J., Baker, A. T., Davey, M. S., Jeeves, M., Sridhar, P., … Willcox, B. E. (2017). BTN3A1 discriminates γδ T cell phosphoantigens from nonantigenic small molecules via a conformational sensor in its B30.2 domain. *ACS Chemical Biology*, *12*(10), 2631–2643.

Salomonsen, J., Skjødt, K., Crone, M., & Simonsen, M. (1987). The chicken erythrocyte-specific MHC antigen. Characterization and purification of the B-G antigen by monoclonal antibodies. *Immunogenetics*, *25*(6), 373–382.

Salomonsen, J., Dunon, D., Skjodt, K., Thorpe, D., Vainio, O., Kaufman, J., … Kaufman, J. (1991). Chicken major histocompatibility complex-encoded B-G antigens are found on many cell types that are important for the immune system. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(4), 1359–1363.

Salomonsen, J., Marston, D., Avila, D., Bumstead, N., Johansson, B., Juul-Madsen, H., … Kaufman, J. (2003). The properties of the single chicken MHC classical class II α chain (B-LA) gene indicate an ancient origin for the DR/E-like isotype of class II molecules. *Immunogenetics*, *55*(9), 605–614.

Salomonsen, J., Sørensen, M. R., Marston, D. A., Rogers, S. L., Collen, T., van Hateren, A., … Kaufman, J. (2005). Two CD1 genes map to the chicken MHC, indicating that CD1 genes are ancient and likely to have been present in the primordial MHC. *Proceedings of the National Academy of Sciences*, *102*(24), 8668–8673.

Salomonsen, J., Chattaway, J. A., Chan, A. C. Y., Parker, A., Huguet, S., Marston, D. A., … Kaufman, J. (2014). Sequence of a complete chicken BG haplotype shows dynamic expansion and contraction of two gene lineages with particular expression patterns. *PLoS Genetics*, *10*(6).

Salomonsen, J., Eriksson, H., Skjodt, K., Lundgreen, T., Simonsen, M., & Kaufman, J. (1991). The adjuvant effect of the polymorphic B-G antigens of the chicken major histocompatibility complex analyzed using purified molecules incorporated in liposomes. *European Journal Of Immunology*, *21*(3), 649–658.

Schat, K. A., Taylor, R. L., & Briles, W. E. (1994). Resistance to Marek's Disease in chickens with recombinant haplotypes of the Major Histocompatibility (B) complex. *Poultry Science*, *73*(4), 502–508.

Schierman, L. W., Watanabe, D. H., & Mcbride, R. A. (1977). Increased growth of Rous sarcomas in chickens pretreated with formalinized syngeneic tumor cells. *European Journal of Immunology*, *7*(10), 710–713.

Schierman, L. W., & Mcbride, R. A. (1967). Adjuvant Activity of Erythrocyte Isoantigens. *Science*, *156*(3775), 658-659.

Schierman, L. W., & Nordskog, A. W. (1961). Relationship of blood type to histocompatibility in chickens. *Science*, *134*(3484), 1008–1009.

Schou, T. W., Labouriau, R., Permin, A., Christensen, J. P., Sørensen, P., Cu, H. P., … Juul-Madsen, H. R. (2010). MHC haplotype and susceptibility to experimental infections (Salmonella Enteritidis, Pasteurella multocida or Ascaridia galli) in a commercial and an indigenous chicken breed. *Veterinary Immunology and Immunopathology*, *135*(1), 52–63.

Schulten, E. S., Briles, W. E., & Taylor R. L., J. (2009). Rous sarcoma growth in lines congenic for major histocompatibility (B) complex recombinants1. *Poultry Science*, *88*(8), 1601–1607.

Senseney, H. L., Briles, W. E., Abplanalp, H., & Taylor, R. L. (2000). Allelic complementation between MHC haplotypes B(Q) and B17 increases regression of Rous sarcomas. *Poultry Science*, *79*(12), 1736–1740.

Shaw, I., Powell, T. J., Marston, D. a, Baker, K., van Hateren, A., Riegert, P., … Kaufman, J. (2007). Different evolutionary histories of the two classical class I genes BF1 and BF2

illustrate drift and selection within the stable MHC haplotypes of chickens. *Journal of Immunology (Baltimore, Md. : 1950)*, *178*(9), 5744–5752.

Shiina, T., Briles, W. E., Goto, R. M., Hosomichi, K., Yanagiya, K., Shimizu, S., … Miller, M. M. (2007). Extended gene map reveals tripartite motif, C-type lectin, and Ig superfamily type genes within a subregion of the chicken MHC-B affecting infectious disease. *The Journal of Immunology*, *178*(11), 7162–7172.

Silva, A. P., Hauck, A. R., Zhou, B. H., & Gallardo, R. A. (2017). Understanding immune resistance to infectious bronchitis using major histocompatibility complex chicken lines. *Avian Diseases*, *61*(61), 358–365.

Simonsen, M., Crone, M., Koch, C., & Hála, K. (1982). The MHC haplotypes of the chicken. *Immunogenetics*, *16*(6), 513–532.

Simonsen, M., Hála, K., & Nicolaisen, E. M. (1980). Linkage disequilibrium ofMHC genes in the chicken. *Immunogenetics*, *10*(1–4), 103–112.

Skjødt, K., Koch, C., Crone, M., & Simonsen, M. (1985). Analysis of chickens for recombination within the MHC (B-complex). *Tissue Antigens*, *25*(5), 278–282.

Smith, I. A., Knezevic, B. R., Ammann, J. U., Rhodes, D. A., Aw, D., Palmer, D. B., … Trowsdale, J. (2010). BTN1A1, the mammary gland butyrophilin, and BTN2A2 are both inhibitors of T cell activation. *Journal of Immunology (Baltimore, Md. : 1950)*, *184*(7), 3514–3525.

Spagnolo, P., Sato, H., Grutters, J. C., Renzoni, E. A., Marshall, S. E., Ruven, H. J. T., … Welsh, K. I. (2007). Analysis of BTNL2 genetic polymorphisms in British and Dutch patients with sarcoidosis. *Tissue Antigens*, *70*(3), 219–227.

Stone, H. A. (1975). Use of highly inbred chickens in research. USDA-ARS Technical Bulletin No.1514, Washington, DC.

Suzuki, K., Matsumoto, T., Kobayashi, E., Uenishi, H., Churkina, I., Plastow, G., … Mitsuhashi, T. (2010). Genotypes of chicken major histocompatibility complex B locus associated with regression of Rous sarcoma virus J-strain tumors. Poultry Science, 89(4), 651–657.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*(22), 4673–4680.

Tregaskes, C. A., Harrison, M., Sowa, A. K., van Hateren, A., Hunt, L. G., Vainio, O., & Kaufman, J. (2016). Surface expression, peptide repertoire, and thermostability of chicken class I molecules correlate with peptide transporter specificity. *Proceedings of the National Academy of Sciences*, *113*(3), 692–697.

Vantourout, P., Laing, A., Woodward, M. J., Zlatareva, I., Apolonia, L., Jones, A. W., … Hayday, A. C. (2018). Heteromeric interactions regulate butyrophilin (BTN) and BTN-

like molecules governing γδ T cell biology. *Proceedings of the National Academy of Sciences*. *115*(5): 1039-1044.

Vavassori, S., Kumar, A., Wan, G. S., Ramanjaneyulu, G. S., Cavallari, M., El Daker, S., … De Libero, G. (2013). Butyrophilin 3A1 binds phosphorylated antigens and stimulates human γδ T cells. *Nature Immunology*, *14*(9), 908–916.

Vilhelmová, M., Miggiano, V. C., Pink, J. R. L., Hála, K., & Hart-manová, J. (1977). Analysis of the alloimmune properties of a recombinant genotype in the major histocompatibility complex of the chicken. *European Journal of Immunology*, *7*(10), 674–679.

Walker, B. A., Hunt, L. G., Sowa, A. K., Skjødt, K., Göbel, T. W., Lehner, P. J., & Kaufman, J. (2011). The dominantly expressed class I molecule of the chicken MHC is explained by coevolution with the polymorphic peptide transporter (TAP) genes. *Proceedings of the National Academy of Sciences*, *108*(20), 8396–8401.

Wallny, H.-J., Avila, D., Hunt, L. G., Powell, T. J., Riegert, P., Salomonsen, J., … Kaufman, J. (2006). Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. *Proceedings of the National Academy of Sciences*, *103*(5), 1434–1439.

Ziegler, A., & Pink, R. (1976). Chemical properties of two antigens controlled by the major histocompatibility complex of the chicken. *Journal of Biological Chemistry*, *251*(17), 5391–5396.

Publication

Chen, L., Fakiola, M., Staines, K., Butter, C., & Kaufman, J. (2018). Functional alleles of chicken BG genes, members of the butyrophilin gene family, in peripheral T cells. *Frontiers in Immunology*, *9*, 930.